

U
<
x
(

[illegible]

```

1 FILE REFERENCE: 107196.132
2 CURRENT APPLICATION NUMBER: US/09/248,796A
3 CURRENT FILING DATE: 1999-02-12
4 PRIOR APPLICATION NUMBER: US 60/074,725
5 PRIOR FILING DATE: 1998-02-13
6 PRIOR APPLICATION NUMBER: US 60/096,409
7 PRIOR FILING DATE: 1998-08-13
8 NUMBER OF SEQ ID NOS: 28208
9 SEQ ID NO 3819
10 LENGTH: 1230
11 TYPE: DNA
12 ORGANISM: Candida albicans
13 FEATURE:
14 NAME/KEY: unsure
15 LOCATION: (1194)
16 OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
17 US-09-248-796A-3819
18
19 Query March 89.4%; Score 1161; DB 4; Length 1230;
20 Best Local Similarity 98.2%; Pred. No. 3,3e-305;
21 Matches 1173; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```

Query Match	Local Similarity	89.4%	Score 1161	DB 4	Length 1230
Best Match	Similarity	98.2%	Pred. No. 3.3e-305		
Matches 1173	Conservative	0	Mismatches 21	Indels 0	Gaps 0
Qy	7	AAAGCATTTAGTGCACCTGGAAAAAGCATTTCTTGCTGTGTGATATTTGGTCTTGAGCCA	66		
Dp	34	AAAGCATTTAGTGCACCTGGAAAAAGCATTTCTTGCTGTGTGATATTTGGTCTTGAGCCA	93		
Qy	67	ATTATGATGCTTATGTGACAGCAATTTGTCTATCAGAAATGCAAGTTATATACCAAAA	128		
Dp	94	ATTATGATGCTTATGTGACAGCAATTTGTCTATCAGAAATGCAAGTTATATACCAAAA	153		
Qy	127	GGAACAGTTTGAAGAACTCTGAATCAAAATTTCTTCAACCCAAATTTGCAAAAGGAGAA	186		
Dp	154	GGAACAGTTTGAAGAACTCTGAATCAAAATTTCTTCAACCCAAATTTGCAAAAGGAGAA	213		
Qy	187	TGGAAATATCACATATCATCAATATACAGAGAGCCCAAGAGATTGACAGCATTAAT	248		
Dp	214	TGGAAATATCACATATCATCAATATACAGAGAGCCCAAGAGATTGACAGCATTAAT	273		
Qy	247	CCATTTTGAAGGCACTATATTCATGCTTTTAGCTTATATTCACCGACGAAACATTT	306		
Dp	274	CCATTTTGAAGGCACTATATTCATGCTTTTAGCTTATATTCACCGACGAAACATTT	333		
Qy	307	GATCTTGAATCATCATTTTACTCAACCCCTGGATATCTTACAAGAAGATACGAAACC	366		
Dp	334	GATCTTGAATCATCATTTTACTCAACCCCTGGATATCTTACAAGAAGATACGAAACC	393		
Qy	367	AAAGCATCTCTGAATGAGAGAAAAAATTTCTTTACATCTCTGTCATTTACCGAAGTG	426		
Dp	394	AAAGCATCTCTGAATGAGAGAAAAAATTTCTTTACATCTCTGTCATTTACCGAAGTG	453		
Qy	427	GAAAGACCGGATTAAGTTTCAATCCGACAGATTAAGTGTCAAGTTTGGCAAGTTATTA	486		
Dp	454	GAAAGACCGGATTAAGTTTCAATCCGACAGATTAAGTGTGTGGGCAAGTTATTA	513		
Qy	487	TCCATTTTATCCCAATGTTATACGTACGATTAAGATATTTTGGCAACGTTGCACAG	546		
Dp	514	TCCATTTTATCCCAATGTTATACGTACGATTAAGATATTTTGGCAACGTTGCACAG	573		
Qy	547	ATTGCACATTTGTTATGCCCAAAAAAGATAGAGTCTGGGTTTGTGCACTGCATTT	606		
Dp	574	ATTGCACATTTGTTATGCCCAAAAAAGATAGAGTCTGGATTTGTGTTGCACCTGCATTT	633		
Qy	607	TATGGCTGATTTGTATATAGAAAGTTTACGACACTTTGATTAATATGACGTTTACAGTT	666		
Dp	634	TATGGCTGATTTGTATATAGAAAGTTTACGACACTTTGATTAATATGACGTTTACAGTT	693		
Qy	667	CTAGAAAGTGATCTGAGAGATTTCCCAACAGATTTGAAAAAATTTGATGAAATCTGG	726		
Dp	694	CTAGAAAGTGATCTGAGAGATTTCCCAACAGATTTGAAAAAATTTGATGAAATCTGG	753		
Qy	727	GAAATTCAAACATGAAAGATGTACATTTACATACGGAATCAAGTTATTAATGGGTGACGT	786		
Dp	754	GAAATTCAAACATGAAAGATGTACATTTACATACGGAATCAAGTTATTAATGGGTGACGT	813		


```

RESULT 5
US-09-792-024-348/C
: Sequence 348 Application US/09792024
: Patent No. 6783985
GENERAL INFORMATION:
APPLICANT: Roemer, Terry
APPLICANT: Jiang, Bo
APPLICANT: Boone, Charles
APPLICANT: Bussey, Howard
TITLE OF INVENTION: Gene Disruption Methodologies for Drug
FILE REFERENCE: 10182-004-999
CURRENT APPLICATION NUMBER: US/09/792,024
CURRENT FILING DATE: 2001-02-20
NUMBER OF SEQ ID NOS: 490
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 348
LENGTH: 90
TYPE: DNA

```

649 AATGACGTGTTTCAAGTTCTTGAACCTCATCCAAATGAGTGGT
Gaps 0; Indels 101


```

Query Match      3.4%; Score 44.6; DB 4; Length 1356;
Best Local Similarity 47.6%; Prog No. 0.026;
Matches 131; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

CY      1019 TACACGATTAACGCAAAAATCAGAGCTCCATCTGAACCTGATGCCAAACCCAGTTT 1078
          |||||
DB      545 TCCAGGTCGAAGCAGCCAAACCGAAGCTAAATGCTTGGTGGATGATGACGCTAAC 604

```

```

RESULT 10
US-09-621-976-2813
/ Sequence 2813, Application US/09621976
/ Patent No. 6639083
/ GENERAL INFORMATION:
/ APPLICANT: Dumas Milne Edwards, J.B.
/ APPLICANT: Jobert, S.
/ APPLICANT: Giordano, J.Y.
/ TITLE OF INVENTION: ESTs and Encoded Human Proteins.
/ FILE REFERENCE: GENSET.054PR2
/ CURRENT APPLICATION NUMBER: US/09/621,976
/ CURRENT FILING DATE: 2000-07-21
/ NUMBER OF SEQ ID NOS: 19335
/ SOFTWARE: Patent.pm
/ SEQ ID NO 2813
/ LENGTH: 832
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 235..399
US-09-621-976-2813

```

	Query Match	3.4%;	Score 44.2;	DB 4;	Length 832;
	Best Local Similarity	15.2%;	Pred. No. 0.026;		
	Matches	60;	Conservative 164;	Mismatches 167;	Indels 4; Gaps 1
QY	339	ATATCATTCACAGAGATCTACTGAAACCAACATCTCTGAAATGAGAGAAAAACATTCTT	398		
Db	11	WYACWCTTKWMSYMYWTKWYWKTYRHRKKGAKAMKYMWTWTWYWRAYMMGYKK	70		
QY	399	TTTCCATTCCTCGCATCTTACCCGAGTGGAGAAAAAGCCGATTAGGTTTCATCGGCAGAGTT	458		
Db	71	KAMCRKTKYKKKKKKKKYMMWYMGWRBSYMMWMTRTWGYAYRSMMWYWRHCRKKAYY	130		
QY	459	AGTGTCAAGTTGTTGGCAAGATTATATCCGATTTATCCCAATGTTATCGAGCA	518		
Db	131	RRTTCYSSKGMTWMMKMKAMTTWMMKTYIMATYMMMC---WKSRPASWYTCWY	186		
QY	519	TAAAGATATTTTGCACAACGTTSCACAGATTGCAATGTTAGCCCAAAAAGATAGG	578		
Db	187	WCKAKKMTWKSYSYASARSARCCYSCSWMGMSKYMWRMRGRGATGAKMKAWRAS	246		
QY	579	ATCTGGGTTTGATTTTGCACATGTCATTTATAGTCTGATTTGATATATAGAGATTTCAGCC	638		
Db	247	CMBRKRYKAGSKSYSKMSWMCWTRBSKTYTYAARTGYCYRKGGMGKGRWYASKVM	306		
QY	639	AGCTTGATATAAGAGCGTGTTCAGTTCTAGAAAGATGCTCAGAGAGTTCCCCACAGA	698		
Db	307	WKRMMWCMAMRYYSTGTPTASMMWRMYTMMCKMKKYMABABAKRWYAMWAMWRBCAA	366		
QY	699	GTTGAAAAATTTGATTGAAAGATACCTGGGAATTCA	733		
Db	367	AATATATATTTATAGTACAACTTCTGTACTCTTA	401		

;; CURRENT FILING DATE: 2001-11-21
;; NUMBER OF SEQ ID NOS: 5
;; SOFTWARE: Patent In Ver. 2.1
;; SEQ ID NO: 4
;; LENGTH: 728
;; TYPE: DNA
;; ORGANISM: Gossypium hirsutum
;; FEATURE:
;; NAME/KEY: unsure
;; LOCATION: 715
;; OTHER INFORMATION: n can be any nucleotide
US-09-988-863A-4

Query Match

Best Local Similarity 3.1%; Score 40.4; DB 4; Length 728;
Matches 62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 514 ACGAATTAAGATATTTTGCACAAAGTTGCACGATTTGATGCCCCAAAAAG 573
DB 562 ACAGATCTGATATTTGTCATATGATGCTCAAGGCCACCTGATTTGCCAAGTAA 621
QY 574 ATGAGATCTGGCTTTGATGTTGCACACTGCATTTAAG 611
DB 622 GTTGGCAGTGGCTTTGATGCTGCTGCTAATG 659

RESULT 15

US-08-434-823-1
; Sequence 1, Application US/08434823
; Patent No. 5712248

GENERAL INFORMATION:

APPLICANT: Kalman, Sue S.
APPLICANT: Kiehne, Kristine L.
TITLE OF INVENTION: NOVEL INSECTICIDE PROTEIN AND GENE
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sandoz Agro, Inc.
STREET: 975 California Avenue
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/434,823
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/197,998
FILING DATE: 16-FEB-1994
APPLICATION NUMBER: US 08/102,316
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/004,474
FILING DATE: 14-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/844,302
FILING DATE: 27-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Marcus-Wyner, Lynn
REGISTRATION NUMBER: 34,869
REFERENCE/DOCKET NUMBER: Z-702/CONT3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/354-3588
TELEFAX: 415/857-1125

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 4106 base pairs
TYPE: nucleic acid

;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 296..3826
;; OTHER INFORMATION: /codon_start= 296
US-08-434-823-1

Query Match

Best Local Similarity 3.0%; Score 38.8; DB 1; Length 4106;
Matches 58; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 903 GAAGAAATGAGGGAATGCGTGAATAATGACTGACGACCCAGAGACTTATATTAAAGA 962
DB 2929 GAAGAAATGAGAGACAAACGTGAATAATGGAATGGGAAACAATATTGTTATAAAGA 2988
QY 963 GTTAGATCATTTCTGTGAGCCCTTGACTGT 992
DB 2989 GGCAAAAGATCTGTAGATGCTTTATTGT 3018

RESULT 16

US-08-457-366-1
; Sequence 1, Application US/08457366
; Patent No. 5731194

GENERAL INFORMATION:

APPLICANT: Kalman, Sue S.
APPLICANT: Kiehne, Kristine L.
TITLE OF INVENTION: NOVEL INSECTICIDE PROTEIN AND GENE
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sandoz Agro, Inc.
STREET: 975 California Avenue
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/457,366
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/197,998
FILING DATE: 16-FEB-1994
APPLICATION NUMBER: US 08/102,316
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/004,474
FILING DATE: 14-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/844,302
FILING DATE: 27-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Marcus-Wyner, Lynn
REGISTRATION NUMBER: 34,869
REFERENCE/DOCKET NUMBER: Z-702/CONT3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/354-3588
TELEFAX: 415/857-1125

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 4106 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 296..3826

OTHER INFORMATION: /codon_start= 296
US-08-457-366-1

Query Match 3.0%; Score 38.8; DB 1; Length 4106;
Best Local Similarity 64.4%; Pred. No. 1.6;
Matches 58; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 903 GAAGAAATTGAGGAAATCGTGAATAATACGACTGAGCCGAGACTTAATTAAGA 962
DB 2929 GAAGAAATGAGAGCAAAACGTGAATAATTGAGATGGAACAATAATTGTTATAAGA 2988

QY 963 GTTAATCATTTCTGTTGACCTTTGACTGT 992
DB 2989 GCGAAAGAACTGTGATGATGCTTAATTGT 3018

RESULT 17
US-08-956-171E-191/c

; Sequence 191, Application US/08956171E
; Patent No. 6593114

GENERAL INFORMATION:
APPLICANT: Charles Kunach

Gil H. Choi
Patrick S. Dillon

Craig A. Rosen
Steven C. Barash

Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue

CITY: Rockville
STATE: Maryland

COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996

APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman

REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224

TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 191:

SEQUENCE CHARACTERISTICS:
LENGTH: 14078 base pairs

TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 191:

US-08-956-171E-191

Query Match 3.0%; Score 38.6; DB 4; Length 14078;
Best Local Similarity 55.6%; Pred. No. 3.2;
Matches 74; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 5 CAAGAATTTAGTGACCTGGAAGAATTTCTGCTGAGATATTGTTCTTGAC 64
DB 2401 CATTGATGATGTTGTTAGCTATATGACATTTGTTCTGATGCTTAATGAATCTTGTC 2342

QY 65 CAATTATGATGCTTATGACAGCATTTGTCATCAAGATGATCAGTTAATACCA 124
DB 2341 TCGCTTCAATCAAAATGTTTCATGCTGTATTCACGATACCTGTACAGCATCTCTAT 2282

QY 125 AAGAACCAGTTT 137
DB 2281 ATTTAACAATATT 2269

RESULT 18

US-08-781-986A-191/c
; Sequence 191, Application US/08781986A
; Patent No. 6737248

GENERAL INFORMATION:
APPLICANT: Charles Kunach

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255

CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue
CITY: Rockville

STATE: Maryland
COUNTRY: USA

ZIP: 20850
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A

FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Benson, Bob
REGISTRATION NUMBER: 30,446

REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 191:
SEQUENCE CHARACTERISTICS:

LENGTH: 14078 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear

US-08-781-986A-191

Query Match 3.0%; Score 38.6; DB 4; Length 14078;
Best Local Similarity 55.6%; Pred. No. 3.2;
Matches 74; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 5 CAAGAATTTAGTGACCTGGAAGAATTTCTGCTGAGATATTGTTCTTGAC 64
DB 2401 CATTGATGATGTTGTTAGCTTATATGACATTTGTTCTGATGCTTAATGAATCTTGTC 2342

QY 65 CAATTATGATGCTTATGACAGCATTTGTCATCAAGATGATCAGTTAATACCA 124
DB 2341 TCGCTTCAATCAAAATGTTTCATGCTGTATTCACGATACCTGTACAGCATCTCTAT 2282

QY 125 AAGAACCAGTTT 137
DB 2281 ATTTAACAATATT 2269

RESULT 19

US-09-248-796A-2077/c
; Sequence 2077, Application US/09248796A
; Patent No. 6747137

```
/ GENERAL INFORMATION:
/ APPLICANT: Keith Weinstein et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
/ FILE REFERENCE: 107196.132
/ CURRENT APPLICATION NUMBER: US/09/248,796A
/ PRIOR FILING DATE: 1998-02-13
/ PRIOR APPLICATION NUMBER: US 60/074,725
/ PRIOR FILING DATE: 1998-02-13
/ PRIOR APPLICATION NUMBER: US 60/096,409
/ NUMBER OF SEQ ID NOS: 28208
/ SEQ ID NO 2077
/ LENGTH: 1455
/ TYPE: DNA
/ ORGANISM: Candida albicans
US-09-248-796A-2077
```

```
Query Match
Best Local Similarity 3.0%; Score 38.4; DB 4; Length 1455;
Matches 99; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
```

```
QY 70 TATGATCTTATGTGACAGCATTTGTCATCAGCAATGCATGAGTTATACACCAAGGA 129
DB 533 TATTTTCTGATTTGTAATTAATTAACACAGAAATAGCCGAGCAATCTAAGAAATGA 474
QY 130 ACCAGTTTGAAGAATCTAGAAATCAAAATTTCTTACCCCAATTTGCAGGAGGAATGG 189
DB 473 ACAATGCAATACCTTTTCTGAAATGTTTAAATCAATTCATTTGACCAATTAATCTC 414
QY 190 GAATATCAATTCATCAATTAATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 249
DB 413 AATATCAATTAATTAATTTCTTCTGATTAATTCAGCGAGGAGTAATTAATTAATTAAT 354
QY 250 TTTTGAAGGCACTATTAAT 269
DB 353 TTTCTGTGGCTCCATCAAT 334
```

RESULT 20

```
US-09-248-796A-216
/ Sequence 216, Application US/09248796A
/ Patent No. 6747137
/ GENERAL INFORMATION:
/ APPLICANT: Keith Weinstein et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
/ FILE REFERENCE: 107196.132
/ CURRENT APPLICATION NUMBER: US/09/248,796A
/ PRIOR FILING DATE: 1999-02-12
/ PRIOR APPLICATION NUMBER: US 60/074,725
/ PRIOR FILING DATE: 1998-02-13
/ PRIOR APPLICATION NUMBER: US 60/096,409
/ NUMBER OF SEQ ID NOS: 28208
/ SEQ ID NO 216
/ LENGTH: 2214
/ TYPE: DNA
/ ORGANISM: Candida albicans
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (179)
/ OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unkno
US-09-248-796A-216
```

```
Query Match
Best Local Similarity 2.9%; Score 37.8; DB 4; Length 2214;
Matches 72; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
```

```
QY 929 AATAGACTGACACCCAGAGCTTATTAAGAGTTAGATCAATCGTTGAGCCCTTGA 988
DB 425 AATAGCAACAGAGGAAAGTTATGCTGATTTGATTAAGCAATATCAACCCCTTGA 484
```

```
QY 989 CTGTTGGATTAAGACATCAGAAAAGGTTTACAGCAATTAACCAAAAATGAGGTTTC 1048
DB 485 CTGATGATTTATTCATTAATTAATAGTAGCTTAGAAAAGCTTAACAAAGATGAGATTG 544
QY 1049 CAATTGAC 1057
DB 545 CAATGAC 553
```

RESULT 21

```
US-09-248-796A-13337
/ Sequence 13337, Application US/09248796A
/ Patent No. 6747137
/ GENERAL INFORMATION:
/ APPLICANT: Keith Weinstein et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
/ FILE REFERENCE: 107196.132
/ CURRENT APPLICATION NUMBER: US/09/248,796A
/ PRIOR FILING DATE: 1998-02-13
/ PRIOR APPLICATION NUMBER: US 60/074,725
/ PRIOR FILING DATE: 1998-02-13
/ PRIOR APPLICATION NUMBER: US 60/096,409
/ NUMBER OF SEQ ID NOS: 28208
/ SEQ ID NO 13337
/ LENGTH: 762
/ TYPE: DNA
/ ORGANISM: Candida albicans
US-09-248-796A-13337
```

```
Query Match
Best Local Similarity 2.9%; Score 37.4; DB 4; Length 762;
Matches 116; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
```

```
QY 684 GAAGTCCCGACAGAGTGAAGAAAATTGTAAGTAATCGGAATTCAGATCAAGAAAG 743
DB 156 GAATACCTTTTACAGTCAAAACCAACGCTAAGCAAAAGAGATTGAATTCAAAG 215
QY 744 ATGATCAATTAATCAATCAAGATCAATTAATTAATGAGTCAAGGTTGCTCAAGAAC 803
DB 216 AGAGCAGATGACCCGATTTGAATTCGCTGTCGCAATCAATCAATGATGACAC 275
QY 804 ACCCAATGATTAACAGATCAATCAATGAGAAAAGGAGGAGGAGGAGGAGGAGGAGG 863
DB 276 ATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 335
QY 864 TGTGATGACAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 923
DB 336 GGTAGTGAATGTTTACCAATTTCTTACGTAACAACTACAAATGAATTCATCAATCA 395
QY 924 TGAATAA 930
DB 396 TCTTAAA 402
```

RESULT 22

```
US-09-601-198-105/c
/ Sequence 105, Application US/09601198
/ Patent No. 6531583
/ GENERAL INFORMATION:
/ APPLICANT: Casseil, Gail H.
/ APPLICANT: Chen, Ellison Y.
/ APPLICANT: Glass, Jennifer S.
/ APPLICANT: Glass, John I.
/ APPLICANT: Heiner, Cheryl R.
/ APPLICANT: Lefkowitz, Elliot
/ TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
/ FILE REFERENCE: UAB-13452/22
/ CURRENT APPLICATION NUMBER: US/09/601,198
/ PRIOR FILING DATE: 2000-12-08
/ PRIOR APPLICATION NUMBER: 60/073,189
```

;; PRIOR FILING DATE: 1998-01-30
;; NUMBER OF SEQ ID NOS: 181
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO: 105
;; LENGTH: 1550
;; TYPE: DNA
;; ORGANISM: Ureaplasma urealyticum
US-09-601-198-105

Query Match 2.9%; Score 37.4; DB 4; Length 1550;
Best Local Similarity 44.0%; Pred. No. 2.4; Indels 0; Gaps 0;
Matches 158; Conservative 0; Mismatches 201;

QY 12 ATTAGTGCATCTGAAAAGCATTTCTGCTGGATATTTGCTCTTGACCAATTTA 71
DB 372 ATTAAACACCTGATTAATAAACAAGCTTGATGTTATTTAGATGTTAAAGAAAC 313
QY 72 TGATCCTTATGTGACAGCATTTGTCATCAGAAATGATGATTAACCAAAAGAAC 131
DB 312 AGATATATGATGATCTTTTATTAATTAATTAAGATGATTCAGAGATTAATGCGGAAAC 253
QY 132 CAGTTTGAAGAATCTAGATCAAAATTTCTCACCCCAATTTGCAAGGAAATGGGA 191
DB 252 TTTTATTAATCAGCTAAAGGATCAATATTTCTATCTGTTGCCAAATCTACATATAT 193
QY 192 ATATCAATATCATCATTAATACAGAGAGCCAGAGAGTTCAAGTCAAGCATTAATCAT 251
DB 192 TTCTAATACATCTTCAAGAAATTTAAAGAAATATGGATTAATCTAATTAATCAATATGA 133
QY 252 TTTAGAGGCAATATATTCATGCTTTTACCTTATATTAATCAAGCAGCAAGCAATTTGATCT 311
DB 132 AAAACAGACATTAATTTGGTCTTTTATCTTATGAGAGCAAAAGTAAAGATGTGG 73
QY 312 TGAATCATCATTTTACTGAGCCCTGATATCATTCACAAGAAAGATCTGAACCAAGA 370
DB 72 ATATTCGTAATAATGATGATCAACGCTCTAAAGAAATTTCTATGCAAAACATGAATTAAGA 14

RESULT 23

US-09-710-279-3904/c
; Sequence 3904; Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 3904
; LENGTH: 3037
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-710-279-3904

Query Match 2.9%; Score 37.4; DB 4; Length 3037;
Best Local Similarity 57.1%; Pred. No. 3.3; Indels 0; Gaps 0;
Matches 68; Conservative 0; Mismatches 51;

QY 733 AAACATGAAAGATGTACATTAACCATGCAAGTCAAGTCAAGTCAAGTCAAGT 792
DB 264 AAAAATGAAAGATGTATTAATTAATGTAATTAACATGATTAATCTCTATAGAGG 205
QY 793 GGCTGAGAAACACCAATTTGGTATACAGATCAATCTCAATGAAAAAGAAAGCCAGA 851
DB 204 GACTTGTTAAATAATTAAGTTTATTAATCAATGATTCGAATGCTCAAAATTAAGCCAGA 146

RESULT 24
US-09-710-279-3434/c
; Sequence 3434; Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 3434
; LENGTH: 4020
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-710-279-3434

Query Match 2.9%; Score 37.4; DB 4; Length 4020;
Best Local Similarity 57.1%; Pred. No. 3.8;
Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 733 AAACATGAAAGATGTACATTAACCATGCAAGTCAAGTCAAGTCAAGTCAAGT 792
DB 3089 AAAAATGAAAGATGTATTAATTAATGTAATTAACATAGATTAATCTCTATAGAGG 3030
QY 793 GGCTGAGAAACACCAAAATTTGGTATCAGAGTCAATGCAAGTCAAGTCAAGTCAAGT 851
DB 3029 GACTTGTTAAATAATTAAGTTTATTAATCAATGATTCGAATGCTCAAAATTAAGCCAGA 2971

RESULT 25

US-07-671-817A-2
; Sequence 2; Application US/07671817A
; Patent No. 542409
; GENERAL INFORMATION:
; APPLICANT: Ely, Susan
; TITLE OF INVENTION: DNA constructs
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cushman, Darby and Cushman
; STREET: Eleventh floor, 1615 L Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-3601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/671,817A
; FILING DATE: 19910401
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8823068.5
; FILING DATE: 30-SEP-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB89/01157
; FILING DATE: 29-SEP-1989
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 2815 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-671-817A-2

Query Match
Best Local Similarity 2.9%; Score 37.2; DB 1; Length 2815;
Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 903 GAAGGAATTGAGGGAATGCGTGAATAATGCACTGAGACCCGAGACTTATATTAAAGA 962
DB 2642 GAAAAAATGAGAGACAAACGTGAAAAATGGAATGGAACAAATATTGTTATATAAGA 2701
QY 963 GTTAGATCATCTTCTGTGAGCCTTTGACTGT 992
DB 2702 GGCAAAAGAACTGTAGATGCTTTATTGT 2731

RESULT 26
US-08-349-867-22
Sequence 22, Application US/08349867
Patent No. 5508264
GENERAL INFORMATION:
APPLICANT: Bradfisch, Gregory A.
APPLICANT: Thompson, Mark
TITLE OF INVENTION: No. 5508264e1 Pesticidal Compositions
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/349,867
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA86
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ. ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 3444 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-349-867-22

Query Match
Best Local Similarity 2.9%; Score 37.2; DB 1; Length 3444;
Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 903 GAAGGAATTGAGGGAATGCGTGAATAATGCACTGAGACCCGAGACTTATATTAAAGA 962
DB 2550 GAAAAAATGAGAGACAAACGTGAAAAATGGAATGGAACAAATATTGTTATATAAGA 2609
QY 963 GTTAGATCATCTTCTGTGAGCCTTTGACTGT 992

DB 2610 GGCAAAAGAACTGTAGATGCTTTATTGT 2639

RESULT 27
US-08-349-867-26
Sequence 26, Application US/08349867
Patent No. 5508264
GENERAL INFORMATION:
APPLICANT: Bradfisch, Gregory A.
APPLICANT: Thompson, Mark
TITLE OF INVENTION: No. 5508264e1 Pesticidal Compositions
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/349,867
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA86
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ. ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 3444 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-349-867-26

Query Match
Best Local Similarity 2.9%; Score 37.2; DB 1; Length 3444;
Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 903 GAAGGAATTGAGGGAATGCGTGAATAATGCACTGAGACCCGAGACTTATATTAAAGA 962
DB 2550 GAAAAAATGAGAGACAAACGTGAAAAATGGAATGGAACAAATATTGTTATATAAGA 2609
QY 963 GTTAGATCATCTTCTGTGAGCCTTTGACTGT 992
DB 2610 GGCAAAAGAACTGTAGATGCTTTATTGT 2639

RESULT 28
US-08-239-476-22
Sequence 22, Application US/08239476
Patent No. 5527883
GENERAL INFORMATION:
APPLICANT: Thompson, Mark
APPLICANT: Schwab, George E.
TITLE OF INVENTION: Improvement of Delta-Endotoxin Expression in
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida

Db 2550 GAAAAATGAGAGACAAACGTGAAAAATGGAATGGGAAACAAATATTGTTATAAGA 2609
QY 963 GTTAGATCATTCCTGTTGAGCCTTGACTGT 992
Db 2610 GGCAGAAAGAAATCTGTAGATGCTTTATTGT 2639

RESULT 31

US-08-598-305A-26
Sequence 26, Application US/08598305A
Patent No. 5827514
GENERAL INFORMATION:
APPLICANT: BRADFISCH, Gregory A.
APPLICANT: THOMPSON, Mark
APPLICANT: SCHWAB, George E.
TITLE OF INVENTION: No. 5827514e1 Pesticidal Compositions
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606-6669
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/598.305A
FILING DATE: 08-FEB-1996
CLASSIFICATION: 514
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/349,867
FILING DATE: 06-DEC-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Sanders, Jay W.
REGISTRATION NUMBER: 39,355
REFERENCE/DOCKET NUMBER: MA86.D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 3444 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-598-305A-26

Query Match 2.9%; Score 37.2; DB 1; Length 3444;
Best Local Similarity 63.3%; Pred. No. 4;
Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 903 GAAGCAATTGAGGAAATCGTGAATAATAGACTCAGACCCAGACTTATATTAAAGA 962
Db 2550 GAAAAATGAGAGACAAACGTGAAAAATGGAATGGGAAACAAATATTGTTATAAGA 2609
QY 963 GTTAGATCATTCCTGTTGAGCCTTGACTGT 992
Db 2610 GGCAGAAAGAAATCTGTAGATGCTTTATTGT 2639

RESULT 32

US-08-639-923A-22
Sequence 22, Application US/08639923A
Patent No. 5840554
GENERAL INFORMATION:
APPLICANT: Thompson, Mark

APPLICANT: Schwab, George E.
TITLE OF INVENTION: Improvement of Delta-Endotoxin Expression in
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESSES:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/639.923A
FILING DATE: 24-APR-1996
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/239,476
FILING DATE: 06-MAY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA83.D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 372-5800
TELEFAX: (904) 375-8100
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 3444 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-639-923A-22

Query Match 2.9%; Score 37.2; DB 2; Length 3444;
Best Local Similarity 63.3%; Pred. No. 4;
Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 903 GAAGCAATTGAGGAAATCGTGAATAATAGACTCAGACCCAGACTTATATTAAAGA 962
Db 2550 GAAAAATGAGAGACAAACGTGAAAAATGGAATGGGAAACAAATATTGTTATAAGA 2609
QY 963 GTTAGATCATTCCTGTTGAGCCTTGACTGT 992
Db 2610 GGCAGAAAGAAATCTGTAGATGCTTTATTGT 2639

RESULT 33

US-08-639-923A-26
Sequence 26, Application US/08639923A
Patent No. 5840554

GENERAL INFORMATION:
APPLICANT: Thompson, Mark
APPLICANT: Schwab, George E.
TITLE OF INVENTION: Improvement of Delta-Endotoxin Expression in
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESSES:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/639,923A
FILING DATE: 24-APR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,476
FILING DATE: 06-MAY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA63.D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 3444 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-639-923A-26

Query Match 2.9%; Score 37.2; DB 2; Length 3444;
Best Local Similarity 63.3%; Pred. No. 4;
Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 903 GAAGAAATTGAGGAAATCGTGAATAATACGACTCAGACCCAGACCTTATATTAAAGA 962
Db 2550 GAAAAAATGAGAGACAAACGTGAATAATTTGAAATCGGAACAAATATTGTTATTAAGA 2609
Qy 963 GTTAGATCATCTCTGTTGAGCCCTTGACTGT 992
Db 2610 GGCAAAAGATCTGTAGATGCTTATTGT 2639

RESULT 34
PCT-US95-05431-22
Sequence 22, Application PC/TUS9505431
GENERAL INFORMATION:
APPLICANT:
APPLICANT: Street address: 5501 Oberlin Drive
APPLICANT: City: San Diego
APPLICANT: State/Province: California
APPLICANT: Country: US
APPLICANT: Postal code/Zip: 92121
APPLICANT: Phone number: (619) 453-8030 Fax number: (619) 453-6991
APPLICANT: Telex number:
TITLE OF INVENTION: Improvement of Delta-Endotoxin Expression in
TITLE OF INVENTION: Pseudomonas fluorescens
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05431
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 3444 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US95-05431-26

REFERENCE/DOCKET NUMBER: MA63
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 3444 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US95-05431-22

Query Match 2.9%; Score 37.2; DB 5; Length 3444;
Best Local Similarity 63.3%; Pred. No. 4;
Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 903 GAAGAAATTGAGGAAATCGTGAATAATACGACTCAGACCCAGACCTTATATTAAAGA 962
Db 2550 GAAAAAATGAGAGACAAACGTGAATAATTTGAAATCGGAACAAATATTGTTATTAAGA 2609
Qy 963 GTTAGATCATCTCTGTTGAGCCCTTGACTGT 992
Db 2610 GGCAAAAGATCTGTAGATGCTTATTGT 2639

RESULT 35
PCT-US95-05431-26
Sequence 26, Application PC/TUS9505431
GENERAL INFORMATION:

APPLICANT:
APPLICANT: Street address: 5501 Oberlin Drive
APPLICANT: City: San Diego
APPLICANT: State/Province: California
APPLICANT: Country: US
APPLICANT: Postal code/Zip: 92121
APPLICANT: Phone number: (619) 453-8030 Fax number: (619) 453-6991
APPLICANT: Telex number:
TITLE OF INVENTION: Improvement of Delta-Endotoxin Expression in
TITLE OF INVENTION: Pseudomonas fluorescens
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05431
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA63
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 3444 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US95-05431-26

```

Query Match      2.9% Score 37.2; DB 5; Length 3444;
Best Local Similarity 63.3%; Pred. No. 4;
Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 903 GAAGAAATTGAGGAATGCCTGGAAAAAATACGACTCAGACCAGAGACTTATTTAAAGA 962
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2550 GAAAATAATGAGAGACAACCGTGAAAAATGTGAATGGGAAACAAATATTGTTATTAAGA 2605
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 963 GTTAGATCATTCCTGTGAGCCCTTGACTGT 992
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2610 GGCAAAAGAAATCTGTAGATGCTTTATTTGT 2639
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 36
US-08-349-867-20
Sequence 20, Application US/08349867
Patent No. 5508264
GENERAL INFORMATION:
APPLICANT: Bradtsch, Gregory A.
APPLICANT: Thompson, Mark E.
APPLICANT: Schwab, George E.
TITLE OF INVENTION: No. 5508264el Pesticidal Compositions
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanhik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/349,867
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanhik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M86
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 3450 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-349-867-20

Query Match      2.9% Score 37.2; DB 1; Length 3450;
Best Local Similarity 63.3%; Pred. No. 4;
Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 903 GAAGAAATTGAGGAATGCCTGGAAAAAATACGACTCAGACCAGAGACTTATTTAAAGA 962
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2556 GAAAATAATGAGAGACAACCGTGAAAAATGTGAATGGGAAACAAATATTGTTATTAAGA 2615
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 963 GTTAGATCATTCCTGTGAGCCCTTGACTGT 992
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2616 GGCAAAAGAAATCTGTAGATGCTTTATTTGT 2645
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 37
US-08-239-476-20
Sequence 20, Application US/08239476
Patent No. 5527883
GENERAL INFORMATION:
```

```

1  APPLICANT: Thompson, Mark E.
2  APPLICANT: Schwab, George E.
3  TITLE OF INVENTION: Improvement of Delta-Endotoxin Expression in
4  TITLE OF INVENTION: Pseudomonas fluorescens
5  NUMBER OF SEQUENCES: 34
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE: David R. Saliwanchik
8  STREET: 2421 N.W. 41st Street, Suite A-1
9  CITY: Gainesville
10 STATE: Florida
11 COUNTRY: USA
12 ZIP: 32606
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: Patent'n Release #1.0, Version #1.25
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/239,476
20 FILING DATE:
21 CLASSIFICATION: 435
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Saliwanchik, David R.
24 REGISTRATION NUMBER: 31,794
25 REFERENCE/DOCKET NUMBER: N483
26 TELECOMMUNICATION INFORMATION:
27 TELEFAX: (904) 372-5800
28 TELEPHONE: (904) 372-8100
29 INFORMATION FOR SEQ ID NO: 20:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 3450 base pairs
32 TYPE: nucleic acid
33 STRANDEDNESS: single
34 TOPOLOGY: linear
35 MOLECULE TYPE: DNA (genomic)
36 US-08-239-476-20
37
38 Query Match 2.9%; Score 37.2; DB 1; Length 3450;
39 Best Local Similarity 63.3%; Pred. No. 4;
40 Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
41
42 QY 903 GAAGATTGAGGGAATGCGTGAATAATCGACTCAGACCCAGAGACTTATATTAAAGA 962
43 Db 2556 GAAAAATGAGAGACACAACGTGAAAAATTCGATGGGAAACAATATGTTTAAAGA 2615
44
45 QY 963 GTTAGATCATTCCTGTTGAGCCCTTTGACTGT 992
46 Db 2616 GGCMAAGAATCTGTAGATGCTTATTGT 2645
47
48 RESULT 38
49 US-08-598-305A-20
50 Sequence 20, Application US/08598305A
51 Patent No. 5827514
52 GENERAL INFORMATION:
53 APPLICANT: BRADFISCH, Gregory A.
54 APPLICANT: THOMPSON, Mark
55 APPLICANT: SCHWAB, George E.
56 TITLE OF INVENTION: No. 5827514el Pesticidal Compositions
57 NUMBER OF SEQUENCES: 38
58 CORRESPONDENCE ADDRESS:
59 ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
60 STREET: 2421 N.W. 41st Street, Suite A-1
61 CITY: Gainesville
62 STATE: FL
63 COUNTRY: USA
64 ZIP: 32606-6669
65 COMPUTER READABLE FORM:
66 MEDIUM TYPE: Floppy disk
67 COMPUTER: IBM PC compatible
68 OPERATING SYSTEM: PC-DOS/MS-DOS
69 SOFTWARE: Patent'n
70 CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/598,305A
FILING DATE: 08-FEB-1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/349,867
FILING DATE: 06-DEC-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Sanders, Jay M.
REGISTRATION NUMBER: 39,355
REFERENCE/DOCKET NUMBER: MA86.D1
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 3450 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-598-305A-20

Query Match 2.9%; Score 37.2; DB 1; Length 3450;
Best Local Similarity 63.3%; Pred. No. 4;
Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 903 GAAGGAATTGAGGAAATCGTGAATAATACGACTCAGCCGAGACTTATTATTAAGA 962
DB 2556 GAAAAAATGAGAGACAAACGTGAAAAATTGAAATGGGAAACAAATATTGTTATTAAGA 2615
QY 963 GTTAGATCATTTCTGTTGAGCCTTTGACTGT 992
DB 2616 GGCAAAAGATCTGTAGATGCTTTATTGT 2645

RESULT 39
US-08-639-923A-20
Sequence 20, Application US/08639923A
Patent No. 5640554
GENERAL INFORMATION:
APPLICANT: Thompson, Mark
APPLICANT: Schwab, George E.
TITLE OF INVENTION: Improvement of Delta-Endotoxin Expression in
TITLE OF INVENTION: Pseudomonas fluorescens
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSER: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/639,923A
FILING DATE: 24-APR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,476
FILING DATE: 06-MAY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA83.D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800

INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 3450 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-639-923A-20

Query Match 2.9%; Score 37.2; DB 2; Length 3450;
Best Local Similarity 63.3%; Pred. No. 4;
Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 903 GAAGGAATTGAGGAAATCGTGAATAATACGACTCAGCCGAGACTTATTATTAAGA 962
DB 2556 GAAAAAATGAGAGACAAACGTGAAAAATTGAAATGGGAAACAAATATTGTTATTAAGA 2615
QY 963 GTTAGATCATTTCTGTTGAGCCTTTGACTGT 992
DB 2616 GGCAAAAGATCTGTAGATGCTTTATTGT 2645

RESULT 40
PCT-US95-05431-20
Sequence 20, Application PC/TUS9505431
GENERAL INFORMATION:

APPLICANT:
APPLICANT: Street address: 5501 Oberlin Drive
APPLICANT: City: San Diego
APPLICANT: State/Province: California
APPLICANT: Country: US
APPLICANT: Postal code/Zip: 92121
APPLICANT: Phone number: (619) 453-8030 Fax number: (619) 453-6991
APPLICANT: Telex number:
TITLE OF INVENTION: Improvement of Delta-Endotoxin Expression in
TITLE OF INVENTION: Pseudomonas fluorescens
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSER: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05431
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA83
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 3450 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US95-05431-20

Query Match 2.9%; Score 37.2; DB 5; Length 3450;
Best Local Similarity 63.3%; Pred. No. 4;
Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

OY	903	GAAGAAATTGAGGGAATGCGTGAATAATAGACTCAGACCCAGAGACTTATATTAAAGA	962
DB	2556	GAATAAATGAGAGACAAACGTGAAAAATTGSAATGGAAACAAATATTGTTATATAAGA	2615
OY	963	GTTAGATCATTCCTGTGTGAGCCTTTGACTGT	992
DB	2616	GGCAAAAGATCTGTAGATGCTTTATTGT	2645

Search completed: January 26, 2005, 06:40:22
 Job time : 96 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: January 26, 2005, 03:39:29 ; Search time 3870 Seconds
(without alignments)
15873.209 Million cell updates/sec

Title: US-10-069-062-6

Perfect score: 1299

Sequence: 1 atgcacaaagatctatgc.....aagactataggtttcataa 1299

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_ov:*
8: gb_ov:*
9: gb_ov:*
10: gb_ov:*
11: gb_ov:*
12: gb_ov:*
13: gb_ov:*
14: gb_ov:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1299	100.0	1299	6	AX087879 Sequence
2	1299	100.0	1299	6	AX087878 Sequence
3	1286.2	99.0	1299	6	AX240804 Sequence
4	1286.2	99.0	1299	6	AX488723 Sequence
5	521	40.1	577	6	AX087876 Sequence
6	470.4	35.2	547	6	AX087874 Sequence
7	323.6	24.9	110000	8	CR380952_03
8	201	15.5	110000	8	CR380952_03
9	158.8	12.2	314422	8	AE016898
10	148.4	11.4	110000	8	CR382123_06
11	132.2	10.2	2399	8	YSCERG8
12	132	10.2	1356	6	BD170980
13	132	10.2	1356	6	BD170980
14	132	10.2	40397	8	SC9959
15	101.6	7.8	110000	8	CR382131_07
16	86.6	6.7	42947	8	SPAC343
17	65	5.0	65	6	AX485723
18	65	5.0	90	6	AX241110
19	56	4.3	963	11	CNS06KLC

20	55.6	4.3	1037	11	CNS06C37	AL401081 T7 end of
21	53.8	4.1	7218	6	166494	166494 Sequence 14
22	51.8	4.0	253050	1	AP000984	AP000984 Sulfolobu
23	50.6	3.9	166918	9	AC009757	AC009757 Homo sapi
24	48.2	3.7	2000	6	AX655393	AX655393 Sequence
25	47.8	3.7	732	1	AY147182	AY147182 Streptoco
26	47.8	3.7	923	1	AY139423	AY139423 Streptoco
27	47.8	3.7	1288	1	SPENM41	X58178 S. pyogenes
28	46.4	3.6	949	1	AY139408	AY139408 Streptoco
29	46.4	3.6	951	1	AY139421	AY139421 Streptoco
30	46.2	3.6	1141	6	AX083744	AX083744 Sequence
31	46	3.5	220761	2	AC101700	AC101700 Mus muscu
32	45.8	3.5	233719	2	AC131407	AC131407 Rattus no
33	45.8	3.5	243143	2	AC125664	AC125664 Rattus no
34	45.6	3.5	1141	6	AX083744	AX083744 Sequence
35	45.4	3.5	581	8	AY279126	AY279126 Arisaema
36	45.4	3.5	105989	2	AC008121	AC008121 Homo sapi
37	45.4	3.5	110000	2	AC009727_2	Continuation (3 of
38	45	3.5	2000	6	AX655393	AX655393 Sequence
39	44.6	3.4	1356	6	AR450463	AR450463 Sequence
40	44.6	3.4	143298	8	AC134240	AC134240 Oryza sat
41	44.4	3.4	726	1	AY138855	AY138855 Streptoco
42	44.2	3.4	832	6	AR415176	AR415176 Sequence
43	44.2	3.4	832	6	AX972010	AX972010 Sequence
44	44.2	3.4	832	6	BD110729	BD110729 EST and e
45	43.8	3.4	96492	5	BX323547	BX323547 Zebrafish

ALIGNMENTS

RESULT 1	AX087879	1299 bp	DNA	linear	PAT 17-MAR-2001
LOCUS	AX087879	Sequence 6 from Patent WO0114533.			
DEFINITION	AX087879				
ACCESSION	AX087879.1	GI:13396872			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
source					
ORIGIN					
Query Match		100.0%; Score 1299; DB 6; Length 1299;			
Best Local Similarity		100.0%; Pred. No. 1.5e-280;			
Matches 1299; Conservative		0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	ATGTCAAAAGCATTTAGTGCACCTGGAAGATTTCTTGCTGTGGATATTGTTCTT	60		
DB	1	ATGTCAAAAGCATTTAGTGCACCTGGAAGATTTCTTGCTGTGGATATTGTTCTT	60		
QY	61	GAGCCAAATTATGATGCTTATGTGACAGATTGTCACGATGATGATGATTAACA	120		
DB	61	GAGCCAAATTATGATGCTTATGTGACAGATTGTCACGATGATGATGATTAACA	120		
QY	121	CCAAAGGACAGTTTGAAGATCTGAATCAAAATTTCTTCAACCCCAATTTGCAAC	180		
DB	121	CCAAAGGACAGTTTGAAGATCTGAATCAAAATTTCTTCAACCCCAATTTGCAAC	180		
QY	181	GAGGATGGGATATCATCATCATCAATCAATCAAGAGAGCCAGAGAGTTGATGATCCG	240		
DB	181	GAGGATGGGATATCATCATCATCAATCAATCAAGAGAGCCAGAGAGTTGATGATCCG	240		

[illegible]

RESULT 2	AX087878	1763 bp	DNA	Linear	PAT 17-MAR-2001
LOCUS	AX087878	Sequence 5 from Patent WO0114533.			
DEFINITION	AX087878				
ACCESSION	AX087878.1	GI:13396871			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
source					
ORIGIN					
Query Match		100.0%;	Score 1299;	DB 6;	Length 1763;
Best Local Similarity		100.0%;	Pred. No. 1.5e-280;		
Matches 1299;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
1	ATGTCAAAGCATTAGTGTGCACTCGAAAGCAATTCCTTGCTGATGATATTTGGTTCTT				60
89	ATGTCAAAGCATTAGTGTGCACTCGAAAGCAATTCCTTGCTGATGATATTTGGTTCTT				148
61	GAGCCAAATTTATGATGCTTATGTGACAGCAATTCATCAGAAATGATCAGTTATPACA				120
149	GAGCCAAATTTATGATGCTTATGTGACAGCAATTCATCAGAAATGATCAGTTATPACA				208
121	CCAAAGGAACCGATTGGAAGAAATCTGAATCTAAATTTCTTCAACCCCAATTTGCAAC				180
209	CCAAAGGAACCGATTGGAAGAAATCTGAATCTAAATTTCTTCAACCCCAATTTGCAAC				268
181	GGAGAAATGGAAATATCATATATCATCAATATACAGAAAGCCGAGAAATTCAGTACGC				240
269	GGAGAAATGGAAATATCATATATCATCAATATACAGAAAGCCGAGAAATTCAGTACGC				328
241	ATAATATCAATTTTAAAGGCACTAATATATCATGTTTACTTATATTTCAACCGACGA				300
329	ATAATATCAATTTTAAAGGCACTAATATATCATGTTTACTTATATTTCAACCGACGA				388
301	GCATTTGATCTTGAATATCATATATCATCAATATACAGAAAGCCGAGAAATTCAGTACGC				360
389	GCATTTGATCTTGAATATCATATATCATCAATATACAGAAAGCCGAGAAATTCAGTACGC				448
361	GAAACCAAGATCTCTGAAATGAGAAAGAAACATTTCTTTACCATTTCTGTCATACC				420
449	GAAACCAAGATCTCTGAAATGAGAAAGAAACATTTCTTTACCATTTCTGTCATACC				508
421	GAGTGGAAAGACCGGATTTAGGTTTCATGAGGAGATTTAGTGTGATTTGTCACCAAGT				480
509	GAGTGGAAAGACCGGATTTAGGTTTCATGAGGAGATTTAGTGTGATTTGTCACCAAGT				568
481	TTATTTATCCCATTTTATCCCAATGTTATCAGTAAGAAATATTTGACACAGTT				540
569	TTATTTATCCCATTTTATCCCAATGTTATCAGTAAGAAATATTTGACACAGTT				628
541	GCAAGATTCACATTTGTTATGCCCCCAAAAAGATGAGATCTGGGTTTGAATGTCACCT				600
629	GCAAGATTCACATTTGTTATGCCCCCAAAAAGATGAGATCTGGGTTTGAATGTCACCT				688
601	GCAATTTATGCTGATTTGATATGAGGATTTGACGACGCTTTGATTAAGAGCTGTTT				660
689	GCAATTTATGCTGATTTGATATGAGGATTTGACGACGCTTTGATTAAGAGAGCTGTTT				748
661	CAGGTTCTAGAAAGATGATCTCGAAGATTCGCCACAGATTTGAAAAATTTGATGAAGT				720

Db 749 CAGGTTCTAGAAAGGATCTCTAGAAAGTCCACAGAGTTGAAAAATTTGATTGAAAGT 808
Qy 721 AACTGGAAATTCAAACATGAAGATGTACATTACCAATAGCAATCAAGTTATTAATGAGT 780
Db 809 AACTGGAAATTCAAACATGAAGATGTACATTACCAATAGCAATCAAGTTATTAATGAGT 868
Qy 781 GACGTCAAGGGTGGCTCAGAAAACCCCAATTTGTATCAAGATCTCAATCCAAATGAAAAAG 840
Db 869 GACGTCAAGGGTGGCTCAGAAAACCCCAATTTGTATCAAGATCTCAATCCAAATGAAAAAG 928
Qy 841 GAAAAAGCCGAAAGAAAGCTCTGTGTGTATGACCACTTAATATGTCCTCAATTACAGTTT 900
Db 929 GAAAAAGCCGAAAGAAAGCTCTGTGTGTATGACCACTTAATATGTCCTCAATTACAGTTT 988
Qy 901 ATGAAGGAATGAGGAAATGCGTGAATAATGACTCAGACCCAGAGCTTAATTAA 960
Db 989 ATGAAGGAATGAGGAAATGCGTGAATAATGACTCAGACCCAGAGCTTAATTAA 1048
Qy 961 GAGTTAGATCATTTCTGTGAGCCTTGTGACTGTGCAATTAGAACAATCAGAAAAAGGTTA 1020
Db 1049 GAGTTAGATCATTTCTGTGAGCCTTGTGACTGTGCAATTAGAACAATCAGAAAAAGGTTA 1108
Qy 1021 CAGCATTTAACAACAAAAATCAGAGTTCCAAATTGAACCTGATGTCAAACCAAGTTGTTG 1080
Db 1109 CAGCATTTAACAACAAAAATCAGAGTTCCAAATTGAACCTGATGTCAAACCAAGTTGTTG 1168
Qy 1081 GACCGTTGCAAGAGATTCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
Db 1169 GACCGTTGCAAGAGATTCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1228
Qy 1141 GATGCAATAGCTGTATAGTGTGTGAAAAATCAAGTGGAAATTTTAAGCAGAAAACTCTT 1200
Db 1229 GATGCAATAGCTGTATAGTGTGTGAAAAATCAAGTGGAAATTTTAAGCAGAAAACTCTT 1288
Qy 1201 GAAAAATCCAGATTTATTTTCAATATGTTTACTGCGTTGATTTGGAAGCAAAACAGAGGT 1260
Db 1289 GAAAAATCCAGATTTATTTTCAATATGTTTACTGCGTTGATTTGGAAGCAAAACAGAGGT 1348
Qy 1261 GTACTTGAAGAAAAACCGAAGACTATATAGTTTATTA 1299
Db 1349 GTACTTGAAGAAAAACCGAAGACTATATAGTTTATTA 1387

RESULT 3
LOCUS AX240804 1299 bp DNA linear PAT 26-SEP-2001
DEFINITION Sequence 42 from Patent WO0160975.
ACCESSION AX240804
VERSION AX240804.1 GI:15797740
KEYWORDS
SOURCE
ORGANISM
Candida albicans
Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE
AUTHORS 1 Roemer T., Jiang B., Boone C. and Bussey H.
TITLE Gene disruption methodologies for drug target discovery
JOURNAL Patent: WO 0160975-A 42 23-AUG-2001;
Elitra Pharmaceuticals, Inc. (US)
FEATURES
source location/Qualifiers
1. 1299
/organism="Candida albicans"
/mol_type="unassigned DNA"
/db_xref="taxon:5476"

ORIGIN

Query Match 99.0%; Score 1286.2; DB 6; Length 1299;
Best Local Similarity 99.4%; Pred. No. 1,1e-277;
Matches 1291; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 1 ATGTCAAAAAGCATTTAGTGCACCTGGAAGAAAGCATTTCTTCTGTGTGATATTTGTTCTT 60
Db 1 ATGTCAAAAAGCATTTAGTGCACCTGGAAGAAAGCATTTCTTCTGTGTGATATTTGTTCTT 60

Db 1 ATGTCAAAAAGCATTTAGTGCACCTGGAAGAAAGCATTTCTTCTGTGTGATATTTGTTCTT 60
Qy 61 GAGCCAAATTTATGATGCTTATGTGACAGATTTGTCACGAATGATGATATTAACA 120
Db 61 GAGCCAAATTTATGATGCTTATGTGACAGATTTGTCACGAATGATGATATTAACA 120
Qy 121 CCAAAAAGAACAGTTTGAAGATCTAGAAATCAAAAATTTCTTCAACCCCAATTTGCAAC 180
Db 121 CCAAAAAGAACAGTTTGAAGATCTAGAAATCAAAAATTTCTTCAACCCCAATTTGCAAC 180
Qy 181 GGAAGATGGGAATATCAATATCAATCAATATCAAGAAAGCCAGAAAGTTGATCAAGC 240
Db 181 GGAAGATGGGAATATCAATATCAATCAATATCAAGAAAGCCAGAAAGTTGATCAAGC 240
Qy 241 ATAAATCCATTTTGAAGCACTATATTCATCGTTTATAGTTTATATTAATTAACCCGAA 300
Db 241 ATAAATCCATTTTGAAGCACTATATTCATCGTTTATAGTTTATATTAATTAACCCGAA 300
Qy 301 GCATTTGATCTTGAATATCATTTTATCTGAGACCTGATATCATTCACAAAGATATCT 360
Db 301 GCATTTGATCTTGAATATCATTTTATCTGAGACCTGATATCATTCACAAAGATATCT 360
Qy 361 GAAACCAAGATCTCTGGAATGGAAGAAAAATTTCTTTACATTTCTGTCGCAATTAAC 420
Db 361 GAAACCAAGATCTCTGGAATGGAAGAAAAATTTCTTTACATTTCTGTCGCAATTAAC 420
Qy 421 GAATGGAAGAAAGCCGATTTAGTTTATCGGAGGATTAAGTCAAGTTGTTGCCAAGT 480
Db 421 GAATGGAAGAAAGCCGATTTAGTTTATCGGAGGATTAAGTCAAGTTGTTGCCAAGT 480
Qy 481 TTATATCCCATTTTATCCCAATGTTATCAGTACGAATTAAGATATTTTGCAACGTT 540
Db 481 TTATATCCCATTTTATCCCAATGTTATCAGTACGAATTAAGATATTTTGCAACGTT 540
Qy 541 GCACAGATTTGCAATTTGTTATGCCCCAAAAAGATAGATCTGCGTTGATGTGCAACT 600
Db 541 GCACAGATTTGCAATTTGTTATGCCCCAAAAAGATAGATCTGCGTTGATGTGCAACT 600
Qy 601 GCAATTTATGCTGATTTGTATATAGAAATTTGACGCACTTTGATTAATGACGTTT 660
Db 601 GCAATTTATGCTGATTTGTATATAGAAATTTGACGCACTTTGATTAATGACGTTT 660
Qy 661 CAGGTTCTAGAAAGTATCTCTGAGAAAGTCCCAAGAGTTGAAAAATTTGATTGAAAGT 720
Db 661 CAGGTTCTAGAAAGTATCTCTGAGAAAGTCCCAAGAGTTGAAAAATTTGATTGAAAGT 720
Qy 721 AACTGGAAATTCAAACATGAAGATGTACATTACCAATGCAATCAAGTTATTAATGAGT 780
Db 721 AACTGGAAATTCAAACATGAAGATGTACATTACCAATGCAATCAAGTTATTAATGAGT 780
Qy 781 GACGTCAAGGGTGGCTCAGAAAACCCCAATTTGTATCAAGATCTCAATGGAAGAAAG 840
Db 781 GACGTCAAGGGTGGCTCAGAAAACCCCAATTTGTATCAAGATCTCAATGGAAGAAAG 840
Qy 841 GAAAAAGCCGAAAGAAAGCTCTGTGTGTATGACCACTTAATATAGTCCCAATTTACAGTT 900
Db 841 GAAAAAGCCGAAAGAAAGCTCTGTGTGTATGACCACTTAATATAGTCCCAATTTACAGTT 900
Qy 901 ATGAAGGAATGAGGAAATGCGTGAATAATGACTCAGACCCAGAGCTTAATTAA 960
Db 901 ATGAAGGAATGAGGAAATGCGTGAATAATGACTCAGACCCAGAGCTTAATTAA 960
Qy 961 GAGTTAGATCATTTCTGTGAGCCTTGTGACTGTGGAATTAAGAATCAGAAAAAGGTTA 1020
Db 961 GAGTTAGATCATTTCTGTGAGCCTTGTGACTGTGGAATTAAGAATCAGAAAAAGGTTA 1020
Qy 1021 CAGCATTTAACAACAAAAATCAGAGTTCCAAATTGAACCTGATGTCAAACCAAGTTGTTG 1080
Db 1021 CAGCATTTAACAACAAAAATCAGAGTTCCAAATTGAACCTGATGTCAAACCAAGTTGTTG 1080
Qy 1081 GACCGTTGCAAGAGATTCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
Db 1081 GACCGTTGCAAGAGATTCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140

LOCUS	AX488723	1299 bp	DNA	linear	PAT 16-AUG-2002
DEFINITION	Sequence 6023 from Patent WO02053728.				
ACCESSION	AX488723				
VERSION	AX488723.1	GI:22322735			
KEYWORDS					

REFERENCE

1. saccharomycetales; mitosporic Saccharomycetales; Candida.

TITLE Chung, B., Boone, C., Bussey, H. and Ohlsen, K.L.
Gene disruption methodologies for drug target discovery
PATENT: WO 02053728-A 6023 11-III-2002.

FEATURES	Location/Qualifiers
source	1 1200

```

/organism="Candida albicans"
/mol_type="unassigned DNA"
/db_xref="taxon:5476"

```

Query Match	99.0%	Score 1286.2;	DB 6;	Length 1299;
Best Local Similarity	99.4%	Pred. No. 1.1e-277;		
Matches 1291; Conservative	0;	Mismatches 8;	Indels 0;	Gaps 0

1. **NAME** _____

d

1 ATGTCAAAAAGCA TTTTACATCCCA CCGTGGA A GGGCTTCTT 60

61 GACCAATTATGAGCTTATGTGACAGATTGTCATCAGCAATGATGATGCACTTATATCA 120

61 GAGCCAATTATGATGCTTATGTGACGCAATTGTCATCAGCAATGCATGCAGTTATAACA 120

131 CCAAAACCAATGATTTTCTTCAACCAATTGCAAC 180

NY

181 GGAGATGGCAATACATAATCATCTAAATTGCGTTCGTTCCGTGCCTC
-----CCCCCCCCTTGCAACC 180

181 GGAGATGGGATATCATACATATCAATACAGAAAACCCAAAGATTCACTACCT 240

241 AAAAAATCCATTTTGTAGAGGCCAATAATTATCGTTTGTAGCTTATATTCACCGACCGAA 300

301 GCAATTGATTCCTTGGAAATTTAAGTCTTAATTCACCGACCGAA 300

301 GCATTGATCTGAATCTTTTACTGGACCCCTCAATTTCAGTAACT
360

361 GAAACCAAGACATCTCGAATGAGAAAAACATTTCTTACCAITTCGTGCAATTACC 420

581 GAAACCAAGACATCCCTCGAATGAGAAAAACTTTCTTTACCATTTCTCGTGCATTACC 420

[illegible]

-----CCACACAGI 480

[illegible]

FEATURES

Location/Qualifiers
 1..577
 /organism="Candida albicans"
 /mol_type="unassigned DNA"
 /db_xref="taxon:5476"

ORIGIN

Query Match 40.1%; Score 521; DB 6; Length 577;
 Best Local Similarity 97.1%; Pred. No. 3.5e-106;
 Matches 561; Conservative 0; Mismatches 13; Indels 4; Gaps 3;

177 AACCGAGAAATGGGAATATCATATATCAATAACAGAAACCCAGAGAAAGTTCAATC 236
 577 AACCGGGAATGGGANTATCATATATCAATTAACGAAA-CCCAAGAAAGTACAGTC 519
 237 ACGCAATAAT-CCATTTTAAAGGCAACATATTCATG- -TTTAAAGTTAATTCACCC 293
 518 ACGCAATAATCCATTTTAAAGGCAACATATTCATGCGTTTAAAGTTAATTCACCC 459
 294 GACCGAAGCAATTTGATCTTGAATATCATATTAATCAGACCTCGATATCATTCACAAGA 353
 458 GACCGAAGCAATTTGATCTTGAATATCATATTAATCAGACCTCGATATCATTCACAAGA 399
 354 AGATATCTGAACCAAGACATCTCGAATGAGAGAAAAACATTTCTTTACCATTCCTGTC 413
 398 AGATATCTGAACCAAGACATCTCGAATGAGAGAAAAACATTTCTTTACCATTCCTGTC 339
 414 CATTAACCAAGGAGAAAGACCGGATTTGATTCATCGGAGATTAAGTCAAGTTGTC 473
 338 CATTAACCAAGGAGAAAGACCGGATTTGATTCATCGGAGATTAAGTCAAGTTGTC 279
 474 CACAAGTTAATATCCATTTTATCCCAATGTTATCAGTACGAATTAAGATTTTGA 533
 278 CACAAGTTAATATCCATTTTATCCCAATGTTATCAGTACGAATTAAGATTTTGA 219
 534 CAACGTTGCAAGATTCGATTTGATTCGCAAAAAAAGATGGAATCTGGGTTGATG 593
 218 CAACGTTGCAAGATTCGATTTGATTCGCAAAAAAAGATGGAATCTGGGTTGATG 159
 594 TGCAACTGCAATTTATGCTGATATGATATGAAAGATTTGACGCCAGCTTTGATTAATGA 653
 158 TGCAACTGCAATTTATGCTGATATGATATGAAAGATTTGACGCCAGCTTTGATTAATGA 99
 654 CGTGTTCGAGTTCTAGAAAGTATCTGAGAAAGTTCCCAAGATTTGAAATTTGAT 713
 98 CGTGTTCGAGTTCTAGAAAGTATCTGAGAAAGTTCCCAAGATTTGAAATTTGAT 39
 714 TGAAGTAACTGGAATTCAAACATGAAGATGTAAT 751
 38 TGCAAGTAACTGGAATTCAAACATGAAGATGTAAT 1

RESULT 6
 AX087874 547 bp DNA linear PAT 17-MAR-2001
 LOCUS AX087874
 DEFINITION Sequence 1 from Patent WO0114533.
 ACCESSION AX087874
 VERSION AX087874.1 GI:13396867

KEYWORDS

ORGANISM
 SOURCE
 Candida albicans
 Candida albicans
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE

1 Rosemond, J.D. and Schnell, N.F.
 Phosphomevalonate kinase (pmk) gene (erg8) from candida albicans
 Patent: WO 0114533-A 1 01-MAR-2001;
 Abstrazenece AB (SP)

FEATURES

Location/Qualifiers
 1..547
 /organism="Candida albicans"
 /mol_type="unassigned DNA"
 /db_xref="taxon:5476"

ORIGIN

Query Match 36.2%; Score 470.4; DB 6; Length 547;
 Best Local Similarity 99.8%; Pred. No. 7.6e-95;
 Matches 471; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

828 CCAATGAAAAAGAAAAAGCAAGAAAGCTCTGTTGTATGACCAAGCTTAATAGTCC 887
 1 CCAATGAAAAAGAAAAAGCAAGAAAGCTCTGTTGTATGACCAAGCTTAATAGTCC 60
 888 CAATTAACGTTAATGAAGAAATGAGGAAATGGCTGAAAAATCGACTGACCCAGA 947
 61 CAATTAACGTTAATGAAGAAATGAGGAAATGGCTGAAAAATCGACTGACCCAGA 120
 948 GACTATATTAAGAGTTAGATCATTCCTGTTGAGCCTTGAAGCTTGGCAATTAAGAAAT 1007
 121 GACTATATTAAGAGTTAGATCATTCCTGTTGAGCCTTGAAGCTTGGCAATTAAGAAAT 180
 1008 CAGAAAGGTTTACAGCAATTAACCAAAATCAGAGGTTCCATTTGAACCTGATGTCCA 1067
 181 CAGAAAGGTTTACAGCAATTAACCAAAATCAGAGGTTCCATTTGAACCTGATGTCCA 240
 1068 AACCCAGTTGTTGACCGTTGTCAGAGATTCCTGTTGTGTTGGTGGTGGTCCAG 1127
 241 AACCCAGTTGTTGACCGTTGTCAGAGATTCCTGTTGTGTTGGTGGTGGTCCAG 300
 1128 TGCTGTGATTCAGATGCAATGATGTTAGTGTGGAATAATCAAGTGGAAATTTTAA 1187
 301 TGCTGTGATTCAGATGCAATGATGTTAGTGTGGAATAATCAAGTGGAAATTTTAA 360
 1188 GCAGAAACCTCTTGAATTCAGATTAATTTTCAATATGTTTACTGGGTTGATTTGAGAA 1247
 361 GCAGAAACCTCTTGAATTCAGATTAATTTTCAATATGTTTACTGGGTTGATTTGAGAA 420
 1248 GCAGAAACCTCTTGAATTCAGATTAATTTTCAATATGTTTACTGGGTTGATTTGAGAA 1299
 421 GCAGAAACCTCTTGAATTCAGATTAATTTTCAATATGTTTACTGGGTTGATTTGAGAA 472

RESULT 7
 CR382139_14/c
 WPCOMMENT
 Sequence split into 21 fragments LOCUS CR382139 Accession CR382139

Fragment Name	Begin	End
CR382139_00	1	110000
CR382139_01	100001	210000
CR382139_02	200001	310000
CR382139_03	300001	410000
CR382139_04	400001	510000
CR382139_05	500001	610000
CR382139_06	600001	710000
CR382139_07	700001	810000
CR382139_08	800001	910000
CR382139_09	900001	1010000
CR382139_10	1000001	1110000
CR382139_11	1100001	1210000
CR382139_12	1200001	1310000
CR382139_13	1300001	1410000
CR382139_14	1400001	1510000
CR382139_15	1500001	1610000
CR382139_16	1600001	1710000
CR382139_17	1700001	1810000
CR382139_18	1800001	1910000
CR382139_19	1900001	2010000
CR382139_20	2000001	2051428

Continuation (15 of 21) of CR382139 from base 1400001 (CR382139 Debaryomyces hanseii c)

Query Match 24.9%; Score 323.6; DB 8; Length 110000;
 Best Local Similarity 56.1%; Pred. No. 3.7e-62;
 Matches 778; Conservative 0; Mismatches 514; Indels 96; Gaps 5;

5 CAAAAGCAATTAAGTCAACCTGGAAGAAAGATTTCTTGTGTGATATTTGTTTGAAC 64

```

Db      87295 CTAGAGCATTCAGTCTCTTGGAAGAAGCATTTTGGGGGTGGAATACCTTGATTAAGAC 87236
Qy      65 CAATTTATGATGCTTATGTGACAGCATTTGATCAGAAATGCAATGATTAATACCAA 124
Db      87235 CAATATATCAATGCTTATGTGACAGCGCTTTCTTCAAGAAAGCAATGCGATTTGTAATAAATA 87176
Qy      125 AAGGAACCAAGT---TGAAAGAAATCTAGATCAAAATTTCTTCAACCCCAATTTGCAAG 181
Db      87175 GAGATTCAGATGAAGTCCATCGTCAACATTCATTAAGCTCCCCCAATTCGAAGAG 87116
Qy      182 GAGATTCGGAATATCAATAT-----TCATCAAAATCAGAGAACCCAGAGAGTTTCACT 235
Db      87115 GTGAATGGAATACGAGATGATTAAGAAATTTAGTAAAGATTCCTACAGAGTGAAC 87056
Qy      236 CAGGCAATTAATCCATTTTATAGAGCAATTAATTCATCGTTTACCTTAATTCACCGCA 295
Db      87055 AAGGTATATATCCGTTTATAGAGGCTACGGTTTTCAGTTTTCATTAATTCACCA 86996
Qy      296 CCGAGCATTTGATCTTGAATATCATCATTTACTCAGACCTCGATATCATTCACAAGAG 355
Db      86995 AAGAAATAATTCATCTTGAAATACCATTTTTCAGACCCAGGCTATCATTCACACATG 86936
Qy      356 ATACTGAATCCAGACATCTTCGAATGAGAAAAAATTTCTTTTACCATCTTCGCCA 415
Db      86935 ACACGACGACTAAATATCGTCAACAGACAAAGAAATTTTGTACCATTCGAACCA 86876
Qy      416 TTACCGAAGTGAAGAACCGGATTAAGTTTCACTCGCAGATTAAGTGTAGTGTGCCA 475
Db      86875 TTAATGCGGTGGCCAAACGATTAAGTTCTCTCGGTGGTGTAGTAGTAACATA 86816
Qy      476 CAAGTTTATATCCCATTTATCCCAATGTTATCAGTACGAATTAAGATTTTGGACA 535
Db      86815 CCGGTTGATATCTGCTTTTGAAGACGTAATAGTAACTTGCAAAATGTAATCACA 86756
Qy      536 AAGTTGACAGATTGCAATTTGTTATGCCCCAAAAAAGATAGATCTGGGTTTGTATG 595
Db      86755 ATGTCTCCAAATAGCCACGCTATGCGACGAGAAAGAAATTTGGTTTGTAGTTG 86695
Qy      596 CAATCGCAATTTATGCTGATTTGATTAATGAAGATTTTACGACGCTTGTATTAATGAC 655
Db      86695 CTACAGAGATCTATGGGTGATTAATATATGAGATTCATCTGCTGTGATTAATGAT 86636
Qy      656 TGTTCAGTTCTTAAGAAAGATCT-----GAGAAGTTCGCCA 694
Db      86635 TGTTCACACGCAATTTTAACTCTCAAAATGATGAACAGTTGAAGCAGACTACGCCA 86576
Qy      695 CAAAGTTGAAAAATGATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 754
Db      86575 CTGCTTAAGTCACTTAAGTGAAGTCCAAATGGGAATTAATCACTAAGTACTTAC 86516
Qy      755 CATACGGAATCAAGATTAATGCGGTGACGTCAAGGTGCTCAGAAAACCACTAATTTG 814
Db      86515 CTCAGGCAATTAAGATTAATGAGTGAATCAAGGCGGCTCCGAAACACTAAGTTGG 86456
Qy      815 TATCAGAGTACTCAATGAGAAAAAAGAAAGCCAGAGAAAGCTCTGTGTGTATGACC 874
Db      86455 TGTCCAGATGTTGAAGTGAAGTGAAGTGAAGTGTCTGCAAGCTCTGAATTAACAAC 86396
Qy      875 AGCTTAATAGTGCATTTAAGTTTATGAAGAAATGAGGAATGCGTGAATAATAG 934
Db      86395 GTCTAACAACAGCAATCTGCACTTCAATGATGCTTATCAGTTGTCATCGCTTCTA 86336
Qy      935 ACTCAGACCCAGAGCTTAATTAAGATTAAGTGAATCT-----973
Db      86335 CTGAACACAGCAACCTACAAAGATTTTCTGCAATTTACAGTCTACAGCGTGCCTG 86276
Qy      974 -----CTGTGAGCTTTGACTG 991
Db      86275 TACTTGAATCAATCTGACGAAACCGACCCCAATTTGCCCCATTCGCAATTTGATG 86216
Qy      992 TTGCGATTAAGAAATCATCAGAAAGAGGTTTACAGCATTTAACAACAAATCAGAGTTCCA 1051
Db      86215 AAGGCATCAAAAATATATAGATCAAAATTTGAGAGATCTAACCGCTTACTCGGTGAGAAA 86156

```

```

Qy      1052 TTGAACCTGATGTCCAAAACCCAGATTGTGGACCGTTGCAAGAGATTCCTGTTGTG 1111
Db      86155 TAGAACCCCAATCCAGACTAATCTTTTGATATCTGATACAGATTAAGGGGTCTCG 86096
Qy      1112 GTGTGTGTTTCCAGGTGCTGTGTGATACGATGCAATAGCTGTATTAAGTTTGGAAAAATC 1171
Db      86095 GTGGATGTTTCCAGGTGCAAGGGGCTATGATGCTATATGCTATTTGGTCAATCGAAGT 86036
Qy      1172 AAGTGGAAATTTTAAGCAAAAACCTTGAATAATCCAGATTATTTTCAATATTTACT 1231
Db      86035 CGATCAATGAATTTTGTGATTCAC---CAACAGACAAATTAAGTTGAGATGTCAT 85979
Qy      1232 GGGTGAATTTGGAAGAGAAACAGAGGTGATCTTGAAGAAAAACAGAAAGCTATATAG 1291
Db      85978 GGTGAATTTGACAGAAAGATATGATGTATTTAGAAAGCAAGTAAATGACTATAGTG 85919
Qy      1292 GTTTATTA 1299
Db      85918 GCTTATTA 85911

```

RESULT 8 CR380952_03 WPCOMMENT

Sequence split into 10 fragments LOCUS CR380952 Accession CR380952

Fragment Name	Begin	End
CR380952_00	1	110000
CR380952_01	100001	210000
CR380952_02	200001	310000
CR380952_03	300001	410000
CR380952_04	400001	510000
CR380952_05	500001	610000
CR380952_06	600001	710000
CR380952_07	700001	810000
CR380952_08	800001	910000
CR380952_09	900001	927101

Continuation (4 of 10) of CR380952 from base 300001 (CR380952 Candida glabrata strain CB

Query Match 15.5%; Score 201; DB 8; Length 110000;
Best Local Similarity 51.2%; Pred. No. 1,1e-34;

Matches 688; Conservative 0; Mismatches 595; Indels 60; Gaps 7;

```

Qy      1 ATGTCAAAAGCATTTAGTACGACCTGGAAGAACATTTCTGTGATTAATTTGGTTCT 60
Db      90882 ATGAAAGGCAATACAGTCTCCAGTAAAGTTTACTGTTGGTGTATCTTGTG 90941
Qy      61 GAGCCATTTATGATGCTTATGTAAGACGATTTGATCAGAAATGCAATGCAATGTTG 116
Db      90942 GATCCAAATATGACTCTTACGTTGTAGCTGTATCATCGGAATGCAATGCTGTTG 91001
Qy      117 -AACCAAAAAGAACCGATTGAAGATCTAATCAAAATTTCTTACCCCAATTTG 175
Db      91002 GAAAGGAAGTATACGAACTTTAGTGGCTTCACATTCAGTTTCCAGTCTCAATTT 91061
Qy      176 CAACGGAATGGAAT---ATCAGATTCATCAAAATCAGAGAGCCAGAGAGT 231
Db      91062 AATATAGTGTGGAATTAATGATTAATTTCAAAATGCTTAATGTCAGATCGAAGTT 91121
Qy      232 CAGTCAAGCATTAATTCATTTTATAGGCAACTATATTCATGTTTATGCTTAATTC 291
Db      91122 TCGGGGAATGCAATTCATTTTATAGAGAACTGTGTTAAAGTTTGTAGTTAAG 91181
Qy      292 CCGACGAGCATTTGATCTTGAATCATCATTTTACTGAGCCCGATATCATTCACA 351
Db      91182 CTGTGATGTTAAACGTTTCCAAATGTTATCAAGATTAATTTACTCAGATGCAAGT 91241
Qy      352 GAAGATCTGAACCAAGCATCC-----TCGAATGGAAGAAAAACATTTCTTATCCAT 405
Db      91242 TATCATATGCAAAACATATGATCACTAGATTCATAGCTTGAAGAGCTTCAAGTTCA 91301
Qy      406 TCTGTGCAATTAACGAATGGAAGAACCGGATTAAGTTTATGCGGATTAAGTGTCA 465

```

```

Db      91302 AAGACCGGTATAGCCCAAGTCCAAAAGAGATTGGATCATCTGCTGCTAGTTCG 91361
Qy      466 GTTGTGGCACAAGTTATATCCCAT---TTATCCCAATGTATCATAGCAATAA 522
Db      91362 GTCTGTACACATGCATTAATGAGTTTCTTAAATGACATTCATGCTCCATGACTT 91421
Qy      523 GATATTTTGCACACGTTGCACAGATTTGCATTTGTTATGCCCAAAAAAGATGAGATCT 582
Db      91422 GATCATCATCATATATCATACATATACATTTGTCAGGACAGGCAAGATCGGACAT 91481
Qy      583 GGGTTGATGTTGCAACATCAATTTATGCTGATGTTATATGAAAGATTTGACGCACT 642
Db      91482 GGGTTGATGTTGCAACGCGCTGTTATATGTTCAATTAAGTATACGAAGATTTCAACTGAT 91541
Qy      643 TTGATTAATAGACGTTTTCAGGTTCTAGAAAGTATCCTGAGAAAGTCCCAAGAGTTG 702
Db      91542 TTAATATCTAGACTACTTCTTATCATAGTATACATACAGAGTACAGATTCGAATG 91601
Qy      703 AAAAAATTGATGAAAGTAACTAGGAAATTCAAACATGAAAGATGTACATTACCATACGA 762
Db      91602 AGTAAATTAATAAATCAATGGAATATAAACAATGAATCCATGCTTGGCAGACAGG 91661
Qy      763 ATCAAGTTATTAATGGGTGACGTCAGAGGTGCTCAGAAACCCCAATTGGTATACGA 822
Db      91662 TTGAGACTTGTGTGGGTGATGTAAGGGGTGTTGAGACATGTTGTTGAAAAA 91721
Qy      823 GTACTCCAAATGAAAAAGAAAAAGCAAGAAAGCTGTTGTATGACCACTTAAT 882
Db      91722 GTGAAACTATGTGTACTTAATTAATCCTCAGAGGAAAGATATACAAACATCAAT 91781
Qy      883 AGTGCCAATTTACAGTTTATGAAAGAAATGAGGAAATGCGTAAAAATATACAGCTCAG 942
Db      91782 GAGGCAAAATTTAGATTTATAGATGAGATGCTTCAATTCAGAAATTAAGAAAGCTGAT 91841
Qy      943 CCAAGACTTAATA-----TTAAAGATTAGATCTCTGTGA----- 980
Db      91842 AATGACAAATGCAAAAGTCTCTTGAAGGCAATTAATGCAAAATGCAACAGTTCAAG 91901
Qy      981 -----GCTTTGACTGTGGGATTAAGAAATAGAAAGGGTACAGAGATTAA 1032
Db      91902 GAAATATCTTCACTGAAACCGCAATTCGAAATTAAGGAAATTTCAAGACTAATTAG 91961
Qy      1033 CAAAAATCAGAGGTTCCAAATGGAACCTGATGTCGCAAAACCAAGTTGTCGACCTGTCA 1092
Db      91962 CAAGAGTCTGATGATTAATTAACAGAAAGCTCAACCAACTAGATGCTGTGA 92021
Qy      1093 GAGATTCCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1152
Db      92022 CAGTTAAATGTTGTTATTTGAGGAGTTGTCCTGCTGCTGCTGCTGCTGCTGCTGCT 92081
Qy      1153 GATATAGTTTGGAAAAATCAAGTGGGAAATTTTAAGCAGAAAACTTTGAAATCCAGAT 1212
Db      92082 ATTAATCTCACAAGAAATACAAATTTGAAATGAAAAACAATGCTCAGAA----- 92132
Qy      1213 TATTTTCATATGTTTACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1272
Db      92133 ---TTGATTCGCTTACCTGTTGATTTAAACAAGAAAGCACTTGAAGCTTGA 92189
Qy      1273 AAACCAAGAACTATATAGGTTT 1295
Db      92190 AACCTGAATATATACAGAGATTT 92212

```

```

RESULT 9
AE016898      314422 bp      DNA      linear      PLN 14-APR-2004
DEFINITION   Eremothecium gossypii chromosome V, section 5 of 5 of the complete
ACCESSION    AE016898 AE016818
VERSION      AE016898.1 GI:44983975
KEYWORDS
SOURCE       Eremothecium gossypii (Ashbya gossypii)
ORGANISM     Eremothecium gossypii

```

```

REFERENCE
AUTHORS      Dieckrich,F.S., Voegel,S., Brachat,S., Lerch,A., Gates,K.,
              Steiner,A., Galfney,T.D. and Philippen,P.
TITLE        The Ashbya gossypii genome as a tool for mapping the ancient
              Saccharomyces cerevisiae genome
JOURNAL      Science 304 (5668), 304-307 (2004)
PUBMED       15001715
REFERENCE
AUTHORS      2 (bases 1 to 314422)
              Gates,K., Dieckrich,F.S., Brachat,S., Voegel,S.E., Lerch,A.,
              Philippen,P. and Galfney,T.
TITLE        Direct Subdivision
JOURNAL      Submitted (20-DEC-2002) Department of Ashbya Genomics, Syngenta,
              Research Triangle Park, NC 27709, USA
COMMENT      This is low coverage sequence generated to identify the complete
              set of genes and the gene order on this chromosome. Regions of low
              quality are identified. Before doing extensive work on any gene
              identified here the sequence should be confirmed.
FEATURES
source       1..314422
              /organism="Eremothecium gossypii"
              /mol_type="genomic DNA"
              /strain="ATCC 10895"
              /db_xref="ATCC:10895"
              /db_xref="taxon:33169"
              /chromosome="v"
              /note="synonym: Ashbya gossypii"
              1..33050
              /organism="Eremothecium gossypii"
              /mol_type="genomic DNA"
              /db_xref="taxon:33169"
              /clone="BAC clone BAG1515"
              16499..70444
              /organism="Eremothecium gossypii"
              /mol_type="genomic DNA"
              /db_xref="taxon:33169"
              /clone="BAC clone BAG1624"
              34420..103446
              /organism="Eremothecium gossypii"
              /mol_type="genomic DNA"
              /db_xref="taxon:33169"
              /clone="BAC clone BAG1078"
              108821..151176
              /organism="Eremothecium gossypii"
              /mol_type="genomic DNA"
              /db_xref="taxon:33169"
              /clone="BAC clone BAG1191"
              154249..227554
              /organism="Eremothecium gossypii"
              /mol_type="genomic DNA"
              /db_xref="taxon:33169"
              /clone="BAC clone BAG1362"
              204900..270712
              /organism="Eremothecium gossypii"
              /mol_type="genomic DNA"
              /db_xref="taxon:33169"
              /clone="BAC clone BAG1577"
              244822..312407
              /organism="Eremothecium gossypii"
              /mol_type="genomic DNA"
              /db_xref="taxon:33169"
              /clone="BAC clone BAG1474"
              93
              /note="region of low quality - less than 90 percent
              confidence"
              complement(<252..>1559)
              /locus_tag="AER306C"
              complement(<252..>1559)
              /locus_tag="AER306C"
              /product="AER306Cp"
              complement(<252..>1559)

```

```

unsure
gene
mRNA
CDS

```

[illegible]

Db 85964 CCCCCGGGACAGCTTGTGAGATATTCATTTGAAGTTTCTGTGAGCCGGCTACCACTCC 86023
Qy 355 GATCTGAAACCAAGACATCTCGAATGAGAAAAACATTTCTTACCATTTCTGTCG 414
Db 86024 CAATCTGATAGATTTGAACACCGGATGAGTTCCGTAAGTTTCAAGATTTCAAGAAAAAC 86083
Qy 415 ATTACCGAAGTGAAGAAAGACCGGATTTAGGTTCAATGCGAGATTTAGTGTCACTTTG 474
Db 86084 ATCACCGATGTTCCAAAGACAGGGTGGGGTCACTGACAGGGCTAGTCAACGTTCTCACT 86143
Qy 475 ACAATTTATTTCCCATTTTATCCCAATG-----TATGATAGATTAAGT 525
Db 86144 GCGCTCTGATATCTGTGTTTATCCGAGATGATGTTCTGTAATTAACATTTAGG 86203
Qy 526 ATTTGACAAAGTTGACAGATTTGATGTTATGCCCCAAAAAAGATGAGATCTGG 585
Db 86204 CTGATTCATTAATTTGTCAAGATCGGCACTTGCAGACCCAGAAAGATGAGCACTGT 86263
Qy 586 TTTGATGTTGCACTGCAATTTATGCTCTGATTTATATAGAAATTTTCAGCCAGCTTG 645
Db 86264 TTTGACGTGACATCCGAGATTTTGGCTGATCATCTACAGGGCTTTCGCTGATG 86323
Qy 646 ATTAATGACGTGTTCAAGTTCTAGAAAGTATCTGAGAGTTCCCAAGAGTTGAAA 705
Db 86324 ATTAATGATCTGCTCTATTTG-----GTGCGCGAGATATGCAAGCTGTGCGGAA 86377
Qy 706 AAATGATTAAGATTAATCTGGAATTCAAACATGAAGATTAATTAATTAACGGAATC 765
Db 86378 GTAAATGATTAAGCGGATTTGGATGTCGCCATGACCGAGTCACTTCCCAAAAGCTTA 86437
Qy 766 AAGTTATTAATGATGACGTCAAGGCTGCTCAAGAACACCAAAATGTTATCAGAGTA 825
Db 86438 GCGATCGATGAGATGTCGCGCGGGGTTCAAGAGCGCTGTTGTCAACAGAGTG 86497
Qy 826 CTCGATGAGAAAAAGAAAAAGCCAGAAAGAGTCTGTTGTATGACAGCTTAATAGT 885
Db 86498 CACCAATGTTTCAAGCTGATCCAGAACAGGTAGCAAGATCTACAGAGAGATTAACAAA 86557
Qy 886 GCCAATTAACAGTTATGAGGAATTTGAGGAATGCGTGAATAATAGACTCAGACCA 945
Db 86558 GGAAACATTCATTTATGTCGCGCTTGGAAGAACTCAATAGATTGCAAGCAAGATCTT 86617
Qy 946 GAGCTTATTAATTAAGATT 965
Db 86618 AGCAATATCAATGAATGCT 86637

RESULT 10
CR382123_06/c

Sequence split into 18 fragments LOCUS CR382123 Accession CR382123

Fragment Name	Begin	End
CR382123_00	1	110000
CR382123_01	100001	210000
CR382123_02	200001	310000
CR382123_03	300001	410000
CR382123_04	400001	510000
CR382123_05	500001	610000
CR382123_06	600001	710000
CR382123_07	700001	810000
CR382123_08	800001	910000
CR382123_09	900001	1010000
CR382123_10	1000001	1110000
CR382123_11	1100001	1210000
CR382123_12	1200001	1310000
CR382123_13	1300001	1410000
CR382123_14	1400001	1510000
CR382123_15	1500001	1610000
CR382123_16	1600001	1710000
CR382123_17	1700001	1753957

Continuation (7 of 18) of CR382123 From base 600001 (CR382123 Kluveromyces laetis strain
Query Match 11.4%; Score 148.4; DB 8; Length 110000;

Best Local Similarity 49.7%; Pred. No. 6,6e-23;
Matches 593; Conservative 0; Mismatches 541; Indels 60; Gaps 6;

Qy 9 AGCATTTAGTGCACCTGGAAAAGACATTTCTGTCGAGATTTGTTGTTCTGAGCAAT 68
Db 12974 ATCTTTATGTCCTCAGGTAAGCCCTTCTACAGCGGTTATTAAGTTTAAACCA 12915
Qy 69 TTATGATGCTTATGTAAGACATTTGATCAAGATGATGAGTATTAACCAAAAG 128
Db 12914 ATATGATCTTATGTCGAGGCTATCTCAAGATGATGCTGTTGTCAACCAAGACA 12855
Qy 129 AACAGTTGAAAAGATC-----TGAATCAAAATTTCTACCCCAATTTGC 176
Db 12854 ATTAATTTGTCATATGTCATCAATTAATCTGTAAGTCAGAAATGATCAAGTTTAA 12795
Qy 177 AAAGGAAAGGGAATATCATATCATCAATACAG---AGAGCCCAAGAAAGTTCA 233
Db 12794 TGACGACAGTGGCAGTATTAAGATGATTCAGCTCAATTAACACCCCAAGAAATGCA 12735
Qy 234 GTACGCAATTAATTCATTTTATGAGCACTATATTCATCGTTTATGCTTATTTCAACC 293
Db 12734 AGAAAAACAAGATCCCTTCATGAAAAGTTTATCAACAGTCTTCAACTACTTTCAACC 12675
Qy 294 GACGGAAGCATTTGATCTTGAATCATCATTTTACTCAGACCTGATATCATTCACAGA 353
Db 12674 CGATTTAAACTACAGAGAAACATAGTATGATTAATTAATCTGAGATGCTGCTATCATTC 12615
Qy 354 AGATATGAAACCAAGACATCTCGAATGAGAAAAAACATTTCTTTACATTCCTGTCG 413
Db 12614 CCAGCGGGAACAAACTGAGAAATGACGATGTAATCTTCTTTTCAATAGTAATTC 12555
Qy 414 CATTAACGGAAGTGAAGAACCCGATTAAGTTGTTTATCGGAGGATTAAGTCAAGTTGTTGC 473
Db 12554 TATCATGTAAGTCCCAAACTGAGGTTGGGCTTCAAGCGGATTTGTTACTGTTTATG 12495
Qy 474 CACAAATTAATTAATCCATTTTATCCCAATGTTATCAATGATGAATTAAGATTTTGG- 531
Db 12494 CAGGCAATTTGTTCTGCTTCAACAAATCTTATGTTTATCATCAAGATGATATGAG 12435
Qy 532 -----CACACGTTGACAGATTTGCACTTTGTTATGCCCCAAAAAGATGATCTGG 584
Db 12434 ATTAATTAACCAATTTGTCACAGTGGCACTTGTGAGCAAGAGTAAATGATGATGG 12375
Qy 585 GTTATGTTGCACTGCAATTTATGATCTGATGTTATATAGAGTTTACAGCCAGCTT 644
Db 12374 CTTTATGTTGTCGACAGTGTACGAAACATCTTACAAACAGATTTGATCCGTACT 12315
Qy 645 GATTAATGACGTGTTCAAGTTCTAGAAAGTATCTGAGAAAGTTCCCAAGAGTTGA 704
Db 12314 CATGAGCA-----ATTACATCAACCATTAATGTCAGCATTAATTTGCTTAAGTTCT 12260
Qy 705 AAAATGATGAAAGTAACTGGAATTCAAACATGAAGATGATTAATTAATGAGAT 764
Db 12259 TTGATTTGA-TGAACCGAATTTGGAATAAAGGCTCTGAGTGTCTTCCCTCAAGATT 12201
Qy 765 CAAGTTATTAATGAGTGAAGTGAAGGAGTGTCAAGAACACCAAAATGATGACAGGT 824
Db 12200 GAGGCTAATATGAGGATGTTAATTAATGATCGAGACAAACAACTGTGACAAAAGT 12141
Qy 825 ACTCCAAATGAAAAAGAAAGCCAGAAAGTCTGTTGTGATGACAGCTTAATAG 884
Db 12140 CAATGAATGATTAATGATTAATGATGATGATGATGATGATGATGATGATGATGAT 12081
Qy 885 TCCCAATTTACAGTTTATGAGAAATTAAGGGAATGCGTGAATAATACATCAGACCC 944
Db 12080 AGCGAACAAGAAATTTGCAATCTGCGTGAATCTGCTTATGATGACCAAGCAC 12021
Qy 945 AGAGCTTATAT-----AAAGTTGATCATTC 974
Db 12020 TTGGAATATGAGGCTTGAATGAAACATTAATGATTAAGAAAGAAAGACACACCACT 11961
Qy 975 TGTGAGGCTTGAAGTGTGCAATTAAGAAATCAAGAAAGGTTTCAAGCATTAACACA 1034

Db 11960 ACTGAGAAATAGTTCATTCCGTTCAATACCAATTAGAGAACATTCAGACTCATTACGGA 11901
Oy 1035 AAAATCAGAGGTTCCATTTGAACCTGATGCCAAACCCAGTGTGGACCGTTGTCAGA 1094
Db 11900 ACAATCAGATGCCGATTTGAACCTGAACTCAACATCATCTTTAAATGCTGCAGATCAG 11841
Oy 1095 GATTCTGCTGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 1148
Db 11840 TTTGCCGGGAGTGTAAACCGGAGTTGTACAGAGTGTGAGAGATATGATGCTAT 11787

RESULT 11

YSCER8 2399 bp DNA linear PLN 27-APR-1993
LOCUS YSCER8
DEFINITION S.cerevisiae phosphomevalonate kinase gene, complete cds.
ACCESSION M63648
VERSION M63648.1 GI:553127
KEYWORDS phosphomevalonate kinase.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE

1 (bases 1 to 2399)
Tsay, Y.H. and Robinson, G.M.
Cloning and characterization of ERG8, an essential gene of
Saccharomyces cerevisiae that encodes phosphomevalonate kinase
Mol. Cell. Biol. 11 (2), 620-631 (1991)

AUTHORS

1 (bases 1 to 2399)
Tsay, Y.H. and Robinson, G.M.
Cloning and characterization of ERG8, an essential gene of
Saccharomyces cerevisiae that encodes phosphomevalonate kinase
Mol. Cell. Biol. 11 (2), 620-631 (1991)

JOURNAL

1846667
On Oct 3, 1994 this sequence version replaced gi:171478.
Original source text: S.cerevisiae DNA.

COMMENT

FEATURES

source 1..2399
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
767..2121
/gene="ERG8"
767..774
/gene="ERG8"
847..2121
/gene="ERG8"
/EC_number="2.7.4.21"
/codon_start=1
/product="phosphomevalonate kinase"
/protein_id="AAA34596.1"
/db_xref="GI:171479"
/translation="MSGLRAFSAPGKALLAGYLVDITYEARFVGLSARMNAVHPY
GSLQSDKEFEVRSKQPKDGEWLVIISPKSGFTPVISGSKNPILEKVIANVFTEK
PMMDYCNRLFVIDFSDDAHSOEDSVTEHGNRLSFHSRIEVRKTVGSSAG
LVTVLTPALSPVSDLENNVDKYREVIHNLAVOAHCOAQKIGSGFDVAAGVGSIR
YRRPPALISNLDIGSATYSGKLALHVNEDNITIKSHLPGSLTMMGDIKNSE
TKLVQVKRWYDSHPRESIKIYTELDPHANSRPMDSKIDRLHETHDDYSDQIFESL
ERNDCTQKPEITVEDAVATIRSFRTKESGADIEPPVOTSLIDDCQITIKGYLT
CLIPGAGYDAIVITKQVDVLRQPLMTKD"

ORIGIN

Query Match 10.2%; Score 132.2; DB 8; Length 2399;
Best Local Similarity 50.6%; Pred. No. 4e-19;
Matches 474; Conservative 0; Mismatches 428; Indels 35; Gaps 5;

Oy 3 GTCAAAAGCAATTTAGTGCACCTGGAAAAGCAATTTCTTGCTGGTGATATTTGCTCTTGA 62
Db 855 GTTGAGAGCCTTCACTGCGCCAGGAAAGCGTTACTAGCTGGTGATATTTAGTTTGA 914
Oy 63 GCCAATTTATGATCTTATGACAGCAATTTGTCTATCAGATGATGAGTATTAACACC 122
Db 915 TACAAATATGAAACATTTGATGCTGATTAATCGCAGAAATGATGCTG-TACCCATC 973
Oy 123 AAAAGGAACAGTTTGAAGAAGATCTAGAAATCAAAATTTCTTCAACCCCAATTTGCAAC 182
Db 974 CTTACGGTTCACTTGCAGAGGCTCTGATAGTTTGAAGTCCGCTGCAAAAGTAAACAATTTA 1033

Oy 183 AGAATGGGAATATACATATCATCAATAACAGAGG-----CCAGAGAACTTC 232
Db 1034 AAGATGGGAGTGGCTGATACATATAGTCCAAAATGAGGCTTCATTCCTGTCATAG 1093
Oy 233 AGTCAGCATTAATTCATTTTATAGAGCACTATATTCATCGTTATAGCTTATATCAAC 292
Db 1094 GCGGATCTAAGAACCTTCATTTGAAGAAAGTTATGCTTAACGATTTAGCTATTAAC 1153
Oy 293 CGAACCAAGCATTTGATCTTGAAG-----TCATCATTTATCTGAGACCTGTGATATC 343
Db 1154 CTATCATGACGACACTACTGCAATAGAAACTGTGTCTGTTATGATATTTCTCGATGATG 1213
Oy 344 ATTCACAAGAGATCTGAAACCAAGACATCTCGAATGAGAAAGAAATTTCTTATAC 403
Db 1214 CTTACATCTTCAGAGAGATAGCGTTTACCGAATCGTGTGCAACAGAAATTAAGTTTTC 1273
Oy 404 ATTCTGTCATTTACCGAATGGAAGAACCGGATTAAGTTGATCGGACGATTAAGTGT 463
Db 1274 ATTGCAAGAAATTTGAAGAAAGTTCCCAAAACAGGGCTGGGCTCTCGGACGTTTATGTC 1333
Oy 464 CAGTTGTGCGACAAAGTTTATATCCCATTTTATCCC-----CAATGTTATCAGTA 514
Db 1334 CAGTTTATCTACAGCTTTGGCCCTCTTTTGTATCGACCTGAAATAATATGATGACA 1393
Oy 515 GGAATTAAGATATTTTGGCAACAGCTTGCACAGATTTGATGCGCAAAAAAG 574
Db 1394 AATATAGAGAAATTTATCATATTTAGCACAAAGTTGCTCATTTGCAAGCTCAGGGTAAA 1453
Oy 575 TAGATCTGGGTTGATGATTTGCAACTGCAATTTATGCTGATTTGATATAGAAATTTTC 634
Db 1454 TTGAAAGCGGTTTGAATGATGACGGCGCGACGATATGATCATCATATATAGAAATTTCC 1513
Oy 635 AGCAGCTTTGATTAATGACGCTGTTTCAAGTTCTAGAAAGTATCTGAAAGTTCCCA 694
Db 1514 CACCGCATTAATCTTAATTTGCGAGATATTT-----GGAAGGCTACTTACGCGAGTA 1567
Oy 695 CAGAGTTGAAAAATTTGATTAAGTAATCTGGGAATTCAAACATGAAGATGTACATTTC 754
Db 1568 AACTGGCGCATTTGGTGTATTAAGAAAGCTGAAATTTATGATTTAAAGTACCATTTAC 1627
Oy 755 CATACGAATCAAGTTATTAATGAGTGACGTCAGAGGTGCTCAAGAACCCCAATTTGG 814
Db 1628 CTTGGGATTTAACTTTATGATGAGGCGATTTAAGATGTTTCAGAAACAGTAAATACGG 1687
Oy 815 TATCAGAGTACTCAATGAGAAAAAGAAAGCCAGAAAGAGCTGTTGTATGATGCC 874
Db 1688 TCCAGAAAGTAAAAAATTTGATGATTTGCGATATGCCAGAAAGCTTGAAAAATATATACAG 1747
Oy 875 AGCTTAATAGGCCAATTTTACAGTTTATGAGAAATTT 911
Db 1748 AACTGATCATGCAAAATTTAGATTTATGATGACT 1784

RESULT 12

BD170980 1356 bp DNA linear PAT 17-JAN-2003
LOCUS BD170980
DEFINITION Process for producing prenyl alcohol.
ACCESSION BD170980.1 GI:27876792
VERSION BD170980.1 GI:27876792
KEYWORDS WO 02053746-A/22.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE

1 (bases 1 to 1356)
Oto, C., Obara, S., Muramatsu, M., Nishi, K. and Totsuka, K.
Process for producing prenyl alcohol
Patent: WO 02053746-A 22 11-JUL-2002;

AUTHORS

TOYOAKA MOTOR CORP, CHIKARA OTO, SHUSEI OBARA, MASAYOSHI MURAMATSU,
KIYOHITO NISHI, KAZUHIKO TOTSUKA
OS Saccharomyces cerevisiae (yeast)
PN WO 02053746-A/22
PD 11-JUL-2002

COMMENT

PF 20-DEC-2001 WO 2001JP011214
 PR 28-DEC-2000 JP 00P 403067
 PI CHIKARA OTO, SHUSEI OBATA, MASAYOSHI MURAMATSU, KIYOHICO NISHII,
 KAZUHIKO TOTSUKA
 PC C12N15/52, C12P7/04, C12N1/19, C12N1/21
 CC Process for producing prenyl alcohol
 FH Key
 FT source
 FT location/Qualifiers
 1.1356
 /organism='Saccharomyces cerevisiae (yeast)'.
 /organism='Saccharomyces cerevisiae'
 /mol_type='genomic DNA'
 /db_xref='taxon:4932'

ORIGIN

Query Match 10.2%; Score 132; DB 6; Length 1356;
 Best Local Similarity 50.5%; Pred. No. 4.7e-19;
 Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

3 GTCAAAAGCATTTAGTGCACCTGGAAGCAATTTCTGCTGGTGAATTTGGTTCTTGA 62
 9 GTTGAGAGCTTCAAGTCCCGAGGAAAGCGTTACTAGCTGGTGAATTTAGTTTAA 68
 63 GCCAATTTATGATGCTTATGTAAGCATTTGTCATCAAGATGATGATGATTAACAC 122
 69 TACAAATATGTAAGCATTTGATGCGATTTATCGGCAAGATGATGATGATGATGAT 128
 123 AAAAGAAC-----CAGTTGAAAAGATCTAGAAATCAAAATTTCTTCAACCCCAATTGC 176
 129 TTACGGTTCATTCAGAGGCTGATTAAGTTTAAAGTGGTGAAGTAAACAATTTAA 188
 177 AAAAGAGATGGAATATCATATCATCAATATACAG--AGAGCCGAGAGAGTTCA 233
 189 AGATGGAGAGTGGCTGTACATATAGTCTTAAAGTGGCTTCTTCTTCTTCAATAG 248
 234 GTACAGCATTAATTCATTTTGAAGCACTATATTCATCGTTTGAATTAATTCAC 293
 249 CGGATCTAAGAACCTTTCATTTGAAGTAAAGTATCGTAAAGTATTAAGTATTAAC 308
 294 GACCGAAGATTTGATCTTGAAG-----TCATCATTTACTCGACCCCTGGATATCA 344
 309 TTAACATGACGACTCTGCAATGAAACTTGTTCGTTATGTAATTTCTCGATGATC 368
 345 TTCACAGAGATCTGTAAGAACCAAGCATCTCGAATGAGAGAAAACATTTCTTACCA 404
 369 CTACCATTTCTGAGAGAGTACGCTTACCAATCGTGCACAGAAATTAAGTTTGA 428
 405 TTCTCGTGCATTAACGAGAGTGAAGAACCGGATTAAGTTTCAATCGGACAGATTAGTTC 464
 429 TTCCGACAGAAATGAAAGAGTTCCTCAAAACAGGGCTGGCTCTGGGAGGTTTACTAC 488
 465 AGTTGTGACCAAGTTTATATCCCATTTTATCC-----CAATGTTATCACTAC 515
 489 AGTTTAACTACAGCTTGGCTCTCTTTTGTATGCACTGGAAGAAATATGTAACA 548
 516 GAATTAAGATTTTGAACAAGCTGCAAGATGCAATGTTTATGAGCCCAAAAAGAT 575
 549 ATATAGAGAGATTTATCATATTTTGAACAAGTGTCTCATGTGCAAGCTCAGGTAAT 608
 576 AGCATCTGGGTTTGAATGTTGCAACTGCAATTTATGCTGATTTGTAATAGAGATTGA 635
 609 TGGAGAGCGGTTTGAATGTAAGCGGCGGACATATGATCTATCAAGTATTAAGAAATTC 668
 636 GCCAGCTTGAATTAATGACGTTTCAAGTTCTGAAGATGATCTGAGAAAGTTCCCA 695
 669 ACCCGATTAATCTCTAATTTGCGAGATTT-----GGAAGTCTACTTACCGCGATA 722
 696 AGAGTTGAAAAAATGATGAAGATGAACGAGAAATCAAAATGAAGATGATCAATAC 755
 723 ACTGGGCACTTTGGTTGATGAAGAGATGGAATTAATTAAGATTAAGCAATTAAC 782
 756 ATACGAATCAAGTATTAATGAGTGAAGTCAAGGGTGGCTGAGAAACCCCAATTTGT 815

DB 783 TTCCGATTAATCTTATGATGAGGCGATATTAAGATGTTAGAAACGTAACCTGCT 842
 816 ATCAGCACTACTCCATGAGAAAAAGAGAACGCAAGAAAGCTGTTGTATGACCA 875
 DB 843 CAGAGAGTAAAGAAATTTGTATGATTTCCGATATGCGCAAGAAAGCTGTAATATACAG 902
 876 GCTTAATAGTCCCAATTTACGATTTATGAAGAAAT 911
 DB 903 ACTGATCATGCAAAATTTCAATTTATGATGATGATGATGATGATGATGATGATGAT 938

RESULT 13

BD171098 1356 bp DNA linear PAT 17-JAN-2003
 BD171098
 Process for producing prenyl alcohol.

ACCESSION BD171098
 VERSION BD171098.1 GI:27876910
 KEYWORDS WO 02053747-A/22.
 SOURCE Saccharomyces cerevisiae (baker's yeast)
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE 1 (bases 1 to 1356)
 Oto, C. and Obata, S.
 Process for producing prenyl alcohol
 Patent: WO 02053747-A 22 11-JUL-2002;
 TOYOTA MOTOR CORP, CHIKARA OTO, SHUSEI OBATA

COMMENT

OS Saccharomyces cerevisiae (yeast)
 PN WO 02053747-A/22
 PD 11-JUL-2002
 PF 20-DEC-2001 WO 2001JP011215
 PR 28-DEC-2000 JP 00P 401701, 28-DEC-2000 JP 00P 403067 PR
 18-SEP-2001 JP 01P 282978

PI CHIKARA OTO, SHUSEI OBATA
 PC C12N15/52, C12P7/04, C12N1/19, C12N1/21
 CC Process for producing prenyl alcohol
 FH Key
 FT source
 FT location/Qualifiers
 1.1356
 /organism='Saccharomyces cerevisiae (yeast)'.
 /organism='Saccharomyces cerevisiae'
 /mol_type='genomic DNA'
 /db_xref='taxon:4932'

FEATURES

source

ORIGIN

Query Match 10.2%; Score 132; -DB 6; Length 1356;
 Best Local Similarity 50.5%; Pred. No. 4.7e-19;
 Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

3 GTCAAAAGCATTTAGTGCACCTGGAAGCAATTTCTGCTGGTGAATTTGGTTCTTGA 62
 9 GTTGAGAGCTTCAAGTCCCGAGGAAAGCGTTACTAGCTGGTGAATTTAGTTTAA 68
 63 GCCAATTTATGATGCTTATGTAAGCATTTGTCATCAAGATGATGATGATTAACAC 122
 69 TACAAATATGTAAGCATTTGATGCGATTTATCGGCAAGATGATGATGATGATGAT 128
 123 AAAAGAAC-----CAGTTGAAAAGATCTAGAAATCAAAATTTCTTCAACCCCAATTGC 176
 129 TTACGGTTCATTCAGAGGCTGATTAAGTTTAAAGTGGCTGGAAGTAAACAATTTAA 188
 177 AAAAGAGATGGAATATCATATCATCAATATACAG--AGAGCCGAGAGAGTTCA 233
 189 AGATGGAGAGTGGCTGTACATATAGTCTTAAAGTGGCTTCTTCTTCTTCAATAG 248
 234 GTACAGCATTAATTCATTTTGAAGCAACTATATTCATCGTTTGAATTAATTCAC 293
 249 CGGATCTAAGAACCTTTCATTTGAAGTAAAGTATTCGCTAAGATTAAGTATTAAC 308
 294 GACCGAAGATTTGATCTTGAAG-----TCATCATTTACTCGACCCCTGGATATCA 344

```

Db      309 TAACATGACGACCTACTGCAATAGAAAATTGTCGTTATTGATATTTCTCTGATGATGC 368
Qy      345 TTCAAGAGAGATACCTGAACCAAGACATCTCTGCAATGAGAAAACATTTCTTTACCA 404
Db      369 CTACACATTTCTGAGAGATACCTGTTACCGAACATCGTGGCAACGAAAGATTGAGTTTCA 428
Qy      405 TTCTGTCGCACTTACCGAAGTGGAAAAAGCCGATTAGTTTCATCGGACGAGATTAGTGC 464
Db      429 TTCCACAGAAATGAAAGAGTTCCCAAAACAGGGCTGGCTCTCGGACAGTTTAGTCAC 488
Qy      465 AGTGTGTCACCAAGTTTATTTATCCATTTATCCC-----CAATGTTACGATAC 515
Db      489 AGTTTAACTACACCTTGGCTCTCTTTTGTATCGAAGCTGGAATAATATGATAGCA 548
Qy      516 GAATTAAGATATTTTGGACAACTGGTGCAGATTGCACATTTGATGCCCCAAAAAGAT 575
Db      549 ATATAGAGAGATTATTCATATTTAGCAAGATTGCTCATATGTCAGCTCAGGGTAAAT 608
Qy      576 AGAATCTGGGTTGATTTGCAACTGCAATTTATGCTGATTTATATAGAAATTTCA 635
Db      609 TGGAGCGGGTTGATGATAGCGGCGGACATATGATATATGATATAGAAATTTCC 668
Qy      636 GCCAGTTTGTAAATGACGTTTCAAGTTTCAAGAAAGTCTGAGAAAGTTCCCGAC 695
Db      669 ACCCGATTATCTCTATTTGCGCAATATTT-----GGAAGTCTACTTACCGCAGTAA 722
Qy      696 AGAGTTGAAAAATTTGATTAAGAAATTAAGTGAATTAAGAAAGATGATACATTACC 755
Db      723 ACTGGCGCATTTGTTGATGATGAAGACCTGGAATATTAAGATTAAGAAATACATTACC 782
Qy      756 ATAGGAATCAATTTATATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 815
Db      783 TTGGGATTAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 842
Qy      816 ATCAGCACTACTCAATGAAAAAGAAAAAGCAAGAAAGCTGCTGTTGATGATGATGAT 875
Db      843 CCAGAGGTAAAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 902
Qy      876 GCTTAATAGTGCATTTACAGTTTATGAAAGATTT 911
Db      903 ACTGATCATGCAAAATTTAGATTTATGATGATGAT 938

RESULT 14
SC9959 40397 bp DNA linear PLN 11-AUG-1997
LOCUS S.cerevisiae chromosome XIII cosmid 9959.
DEFINITION 249939.1 GI:887599
VERSION 249939.1 GI:887599
KEYWORDS dihydrofolate reductase; ERG8; MRE11; MRLP44; MTF1;
phosphomevalonate kinase; ribonuclease H; ribosomal protein L44;
RNH1; transfer RNA-Arg; ubiquitin carboxyl-terminal hydrolase.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
1 (bases 16981 to 22170)
REFERENCE 1 Venema,J. and Tollervey,D.
TITLE RRP5 is required for formation of both 18S and 5.8S rRNA in yeast
JOURNAL EMBO J. 15 (20), 5701-5714 (1996)
MEDLINE 97051828
PUBMED 8896463
2 (bases 1 to 40397)
REFERENCE 2 Skelton,J. and Churcher,C.M.
TITLE Unpublished
JOURNAL 3 (bases 1 to 40397)
AUTHORS Barrell,B., Rajandream,M.A. and Walsh,S.V.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-1995) Saccharomyces cerevisiae chromosome XIII
sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridgeshire
CB10 1RQ E-mail: barrell@sanger.ac.uk
NOTES:
All CDS over 100 codons have been analysed. CDS that are completely

```

FEATURES

CDS

overlapped and those that are overlapped by more than 50% of their length by a larger CDS have been omitted from this analysis.
Details of the omitted CDS are available on request. The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons and the calculated codon adaptation index (CAI) is given for each CDS.
Cosmid 9959 overlapped at 5' by cosmid 8261, emb1 entry SC8261X, accessionno.249809 and at the 3' by cosmid 9408, emb1 entry SC9408, accession no.248756.

location/Qualifiers

```

1..40397
  /organism="Saccharomyces cerevisiae"
  /mol_type="genomic DNA"
  /strain="Ab972"
  /db_xref="taxon:4932"
  /chromosome="XIII"
  /map="13R"
  /clone="cosmid 9959"
  /note="YM9959.01, len: > 1051, CAI: 0.15, incomplete unknownorf, overlaps and extends cosmid 8261 orf YM8261.13"
  /codon_start=1
  /product="unknown"
  /protein_id="CA90190.1"
  /db_xref="GI:887600"
  /db_xref="GOA:Q03661"
  /translation="SSSVEIEVEKVSSEKILDSSTKEKELVPLSTDTTINNSSLNMSI
  YSLDDADAISENITDYPMEIKTTPYVISESVYSSTSYSDNTVAMPQVEYTPS
  FMPDPMSLNDVYRKHDILKSTLAALAPFTKDAIEFVAGVTKSLVTSQHTNIF
  FHKETKQVSDLESTENVTVEFENETDERKNOKKPEVANSITDSITBNDTEKTPS
  AINTNTVGTDSCEDIETASNVEMLRYCEKDMNAEWSGDECVKNDGSKTQIS
  FSTDSPNFOESNDNTEFSSTKYKKNLSLDEDESKLETKLAVDKLBDEESDSY
  EODYADPEPNDGSENENIVKTKDPLGIVEPENKVNKVEEETLFFANYSSTNV
  CNKMTHTDVINOEAQNYEAGEKRYIIONTPREAHISIERIDNAIGNMEIPERS
  CVKTHNEV.FERRATTIENYKALLENNTNHDVQACSDSDQVSTABKNVGSAAK
  HNDIRSSSEISVEPLKESDRSNIFSPPIRVIAVAVGVGVAVVAVSFPVKIDV
  MDSDDNDVNDIGYNDIENKNSSTDASVNMKSKSEKSDDEDAVILGGVAAEAKH
  DNGNNSRVNIDITNGAYEEDSEVROQVKENLKSBEPLVEGLQSEHPEKKDH
  SENEEEDITLYGDTANSIHNSAPDDIKRQLKNSIDENYSORLIEIDSRKQNE
  SDEVNTSRERDLTFEKSVEKYGALIEEDTFSELDISIOHPEHNDLDSNNOERIE
  ELSNPEPAELVELIEGPTETLASKKNDDREKRNITSTDLPSPPEDKXEYDTSY
  PYNSNPETAKKAPSPPEYRTFSTPNVPEHEDAIPTATLEKHDNTVTSVD
  RSHLSHSDVDNEPHNNSINIKVNESEEBHOAVDIPVVEVKEDEQEMPSKSVLEQ
  KPSMELINDKSSPENNDSETNEKDKTKAKKSKRKNYNSRRRKRTIEGSSAASNT
  KRRRGHPKPSRGONTHPDYVK"
  1..312
  /note="overlap with cosmid 8261, EMBL SC8261X, 249809,
  positions 27890 to 28201"
  3363..4718
  /gene="ERG8"
  3363..4718
  /gene="ERG8"
  /note="YM9959.02, ERG8 gene, len: 451, CAI: 0.15,
  SW:ERG8_YEAST_P24521, phosphomevalonate kinase; contains
  PS00627_GHMP kinases putative ATP-binding domain, conflict
  at C-terminus due to a frameshift with ERG8_YEAST"
  /codon_start=1
  /product="Erg8p"
  /protein_id="CA90191.1"
  /db_xref="GI:887601"
  /db_xref="GOA:P24521"
  /db_xref="SCD:50004833"
  /db_xref="Swiss-Prot:P24521"
  /translation="MSELRASFAPGKALAGGYLVDTKYEAFFVGLSARMAVNAHPY
  GSLGSDKFEVRSKQFQDGBMLYHISPSGFTIVSISGSKNPTEKVINAVSEYFK
  PMNDYCNRNLFVIDIFSDAYHSGEDSVTERGMRRLSPSHRTIEYPRKGLGSSAG
  LVTVLITLALSPFVSDLENNYDKYREYVHNLAAQVAHCAQKIGSGFPVLAAGSIR
  YRRFPALISNIPDGSATYGSKLALVDDEEDMNITISNHLPSGLTLMWDIKXSE
  TVKLVOVKKNWYDSHPESLKIYELDHANSRFMDGLSKLDRLHETHDYSDQIFESL

```

```
misc_feature
3810..3848
/gene="ERG8"
/note="P500627 GHMP kinases putative ATP-binding domain"
complement(4978..6492)
/note="YW9599.03c, unknown, len: 504, CAI: 0.15, possible
membrane protein"
/codon_start=1
/product="unknown"
/protein_id="CA90192.1"
/db_xref="GI:887602"
/db_xref="GOA:P32829"
/db_xref="SGD:S0004837"
/db_xref="Swiss-Prot:P32829"
/translation="MTSTRTTLYAYOACACITNCPLSAGIIFPPALAKPIILISEGYHE
LDDPDGRLLCTADDLKNTFALSAVTNTMALPVGKILDMYGRVCGIIGSCILF
LASGNFISAKHLVSLMDPYLVGTTLAVAGPVPFISCPQLANSFPGRSCTVALATGS
FDSSDALFLYLYLYONMPTLVNVSFPLVLYVPLAACOLTIMPHSYKTNVHIA
KVAEGLDNGRLIGDGTSGIIPDEQEROSLAIEREDSIISRPOKRSVLETYE
DKLQKSGGIKGVNLGSAVROIKSPWFYMLLPAULMRLNRYPIATRYTOEVLIN
DPDLAKLNSIFDMLPLGAGVSIPIGLLHDITLTLITLFTSTATGVGLIPN
SETWNLIGVLVLRPFYTVSDYSKVFEPDFTGVYGLSCICGIFNNSOILD
KWTHTFMNPPINLTLVLTVPFSLTLPYRSQILPKPNERGLSSNYQTI"
complement(6685..7356)
/note="YW9599.04c, unknown, len: 223, CAI: 0.10, similar
to SW:DYR_SCHPO_P36591, dihydrofolate reductase [29.4%
identity in 218 aa overlap]"
/codon_start=1
/product="unknown"
/protein_id="CA90193.1"
/db_xref="GI:887603"
/db_xref="SGD:S0004835"
/db_xref="Swiss-Prot:P05015"
/translation="MTKYNLMLGLAQSDYFASKTGFRAEMKLGKLYYPYAPNE
PPADVPDLGVIADAPGDGENTGVMLNDPSTGTFPIQTTIDYLVNLRNGP
FAGIVFSGAGVAGIATDFNGLGLTEBQPLFPAVSGFRPOQYOEDLH
PIVPSLHVQGLDITTEPAKQGLYNCTEDSRLLMHSGHFPVNSGFPKVAQM
LQQLT"
7762..9177
/note="YW9599.05, unknown, len: 471, CAI: 0.17, similar to
putative ubiquitin carboxyl-terminal hydrolases, eg
SW:UBPX_HUMAN_P40818, probable ubiquitin
carboxyl-terminal hydrolase [31.5% identity in 352 aa
overlap]; contains P500973 Ubiquitin carboxyl-terminal
hydrolases family 2 signature 2"
/codon_start=1
/product="unknown"
/protein_id="CAA90194.1"
/db_xref="GI:887604"
/db_xref="GOA:P50102"
/db_xref="SGD:S0004836"
/db_xref="Swiss-Prot:P50102"
/translation="MSICPHIOVFONHSEKDGVLTKTCAARYILNHSVPEKFLNTM
KCGTGEINSGATFMCIOGFGCGNHSPLSHSKOIGIPIGNSNGLLPFPKEDY
IGNILINDAILAKYMDVDCYTMVPSMRRGISGLIMSGTCRPSLLOCLINRY
FIRHMSQTHSNCKVRSPDKFSCALDIYELVGLALTKOASSSTINQOCTIY
LITCAWKINQNLAGSQODAHFWQFINQIHQSYVLDPNKEVSRANKQCECIVH
TVFEGSLSEIVPCGONNSKTIIDPDLIDDKDKLTYELDFHKEEDLQKCEIVH
HCGENSTQDAIKQGIHKLPSVLVQLRPFELHNGSRKLDLDFLEPTVIMKNYC
STKEKDSKENGKVPDIYELIYELIYSHKGTVNEGHYACIKISGGWMPFENDSWSSI
SOEVLKROAYLFTIRQVN"
8992..9045
/note="P500973 Ubiquitin carboxyl-terminal hydrolases
family 2 signature 2"
complement(9622..11700)
/gene="MRB11"
complement(9622..11700)
/note="MRB11"
/note="YW9599.06c, MRB11 gene, len: 692, CAI: 0.14,
SW:MR11_YEAST_P32829, required for repair and meiotic
recombination, conflict, this sequence is longer due to a
```

```
frameshift with SW:MR11_YEAST"
/codon_start=1
/product="Mrb1p"
/protein_id="CA90195.1"
/db_xref="GI:887605"
/db_xref="GOA:P32829"
/db_xref="SGD:S0004837"
/db_xref="Swiss-Prot:P32829"
/translation="MDYDPDPIRLILITDNNHGVNENDPIGDSQWKTFRHEVMILAK
NNNVAVVOSGDLFVFNKPSKSLVQVTKTLCGMDGKPELELSDPSQVFNHDEF
TNVYEDENPNISIVPFI SGNHDSASDLSLCPMDILHNLINHPKVIRESDKIV
VPLPQKSTKLALYGLAVRDERLFRFKDGVTFVPTKEGFWFLMCTQHNHTG
HTNTAFLEQFLPDLNVINWGHHECIENLVHNPDKNFVQIPSSSVATSLCEAEQ
PKYVFLIDIKYGEAPKATPIPLETIRTKMSISIQDVPALPHKDXATSKYLIOVE
EMIRANETKOKLADQSGMVAELPPLRLRVDYASPSVQPIDVOYENPRFS
NRFCVRVANGNVVQFYKRSKPVRSKKTNGSTISRDREKLFSSGGELEVOTLV
NDLNLKMLSLPEVGLNABYKPKFDRKBRKALKEFISHIENBVGILSTNEPRTLQ
DAEEMKALIKOVRANSVRPFPPKENDSTNFAFNQNGDSFRSRRVFTGSPDITQS
HYDNESRITTHISQAESKPTSPKPKRVATKKKIPAFSDSTVISAENELGNNDAD
DVIDIDENIIMVSTDEEDASVGLNGRKTKITRPAASTKTASRRKGRASRTPTDI
LGLSLAKKRX"
complement(join(12007..12245,12393..12450))
/gene="MRPL4"
complement(join(12007..12245,12393..12450))
/note="YW9599.07c, MRPL4 gene, (YMR44), len: 98, CAI:

Query Match      10.2%; Score 132; DB 8; Length 40397;
Best Local Similarity 50.5%; Pred. No. 3,4e-19;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

gene
CDS
3 GTCAAGAAGCATTTAGTGCACCTGAAAGCATTTCTTGCTGTGATATTTGTTCTTGA 62
3371 GTTGAAGCCTTCAAGTCCCGGAGAAAGCGTTACTAGCTGTGATATTTGTTTGA 3430
63 GCCAATTAATGATGCTTATGAGACAGATTTGCATCAGATGATGATGATTAACACC 122
3431 TACAATAATGAAAGCATTTGATGCGATTAATGCGCAAGAAATGATGCTGATGCCATCC 3490
123 AAAAGAAC-----CAGTTGAAAAGATCTAGATCAAAATTTCTTCAACCCCAATTGCG 176
3491 TTACGCTTCATTGCAAGGCTCTGATPAAGTTTGAAGTCGCTGAAAAGTAAACAATTGA 3550
177 AAACGAGAATGGGAATATCATATATCATCAATATACAG---AGAACCCGAGAAAGTTCA 233
3551 AAGTGGAGTGGCTGTACATATATAGTCTTAAAGTGGCTTCAATCCGTTTCCATAGG 3610
234 GTACAGCATTAATTCATTTTGAAGCACTATATTCATCGTTTAACTTATATTCACCC 293
3611 CGGATCTAAGAACCTTTCATTTGAAGAAAGTTATCGTAAAGTATTAAGTACTTAAACC 3670
294 GACCGAAGCATTTGATCTTGAAG-----TCATCATTTTCTAGACCCCTGATATCA 344
3671 TTAACATGACGACCTACTGCAATGAAGAACTTGTTCCTTATTTGATTTCTCGATGATGC 3730
345 TTCACAAGAAATACGAAACCAAGCATCTCGAATGAGAAAGAAACATTTCTTATCA 404
3731 CTACATTTCTCGAGAGATAGCTTACGAACATGTGTGCAACAGAGATTAAGTTTGA 3790
405 TTCTGTGCTCATTAACGAAAGTGAAGAAAGCCGAAATTAAGTTGATCGGACAGATTAAGTGC 464
3791 TTGCGACAGAAATGGAAGAAAGTCCCAAAACAGAGGCTGGGCTCTCGGACAGGTTAGTAC 3850
465 AGTTGTGCGACAAAGTTATTTATCCATTTTATCC-----CAATGTTATCAGTAC 515
3851 AGTTTAACTACAGCTTGGCTCTCTTTTGTATCGGACCTGAGAAATTAATGTAGACA 3910
516 GAATTAAGTATTTGACCAAGCTTGACAGATGATGATGATTTGACCAAAAAAAGAT 575
3911 ATATGAGAAATTAATCATATTTGACCAAGATTTGCTCATTTCAAGCTCAGAGTAAAT 3970
576 AGGATCTGGGTTGATGTTGCAAGTGCATTTATGCTGATTTGATATAGAAATTTGA 635
```

```

Db      3971 TGAAGCGGTTTGAATGACGGCGGACAGATATGATCTATACAGATATAGAAATGCC 4030
Qy      636 GCCAGCTTGTATTAATGACGTGTTTCAAGTCTTAGAAGTATCTGTGAAGTTCCAC 695
Db      4031 ACCCCGATTAATCTCTAATTTCCAGATTT-----GGAATGCTTACCTTACGGCAGTAA 4084
Qy      696 AAGTTGAAAAAATGATTGAAAGTAACTGGGATTCAAACATGAAGATGTATACATTAC 755
Db      4085 ACTGGCGCTTTGGTGTATGAAAGACGTGAATTTTCATTAAGTAAACATTAACATTAAC 4144
Qy      756 ATACGATCAAGTATTAATGAGTGTACGTCAAGGCTGCTCAGAAACCCCAATTGCT 815
Db      4145 TTCGGATTAATCTTATGATGAGGAGATTTAAGATGCTTCAAGAAACAGTAAATCTGCT 4204
Qy      816 ATCAGATTAATCTTCAATGAAAAAGAAAGCCAGAAAGAACTCTGTGTGTATGACCA 875
Db      4205 CCAGAGGTTAAAAAATTTGTATGATTCGATATGCGAGAAAGCTTGAATAATATTAACA 4264
Qy      876 GCTTAATATGTCATTTACAGTTATACAGTTATACAGAAATT 911
Db      4265 ACTGATATCAATTAATTTAGATTAATGATGACT 4300

```

RESULT 15
CR382131_07/c
WPCOMMENT

Sequence split into 43 fragments LOCUS CR382131 Accession CR382131

Fragment Name	Begin	End
CR382131_00	1	110000
CR382131_01	100001	210000
CR382131_02	200001	310000
CR382131_03	300001	410000
CR382131_04	400001	510000
CR382131_05	500001	610000
CR382131_06	600001	710000
CR382131_07	700001	810000
CR382131_08	800001	910000
CR382131_09	900001	1010000
CR382131_10	1000001	1110000
CR382131_11	1100001	1210000
CR382131_12	1200001	1310000
CR382131_13	1300001	1410000
CR382131_14	1400001	1510000
CR382131_15	1500001	1610000
CR382131_16	1600001	1710000
CR382131_17	1700001	1810000
CR382131_18	1800001	1910000
CR382131_19	1900001	2010000
CR382131_20	2000001	2110000
CR382131_21	2100001	2210000
CR382131_22	2200001	2310000
CR382131_23	2300001	2410000
CR382131_24	2400001	2510000
CR382131_25	2500001	2610000
CR382131_26	2600001	2710000
CR382131_27	2700001	2810000
CR382131_28	2800001	2910000
CR382131_29	2900001	3010000
CR382131_30	3000001	3110000
CR382131_31	3100001	3210000
CR382131_32	3200001	3310000
CR382131_33	3300001	3410000
CR382131_34	3400001	3510000
CR382131_35	3500001	3610000
CR382131_36	3600001	3710000
CR382131_37	3700001	3810000
CR382131_38	3800001	3910000
CR382131_39	3900001	4010000
CR382131_40	4000001	4110000
CR382131_41	4100001	4210000
CR382131_42	4200001	4224103

Continuation (8 of 43) of CR382131 from base 700001 (CR382131 Yarrowia lipolytica chromo

Query Match 7.8%; Score 101.6; DB 8; Length 110000;
Best Local Similarity 49.7%; Pred. No. 28-12;
Matches 319; Conservative 0; Mismatches 314; Indels 9; Gaps 2;

```

Qy      394 TTCTTACCATTCCTCGTCCATTCACGAAGTGAAGAAAGACCGATTGTTTCATGCGCA 453
Db      1668 TTTCATACCAAAAAGGCGATTCACAGAGGCTCTTAAAGGGGCTGTGAGTCCGCT 16609
Qy      454 GGATTAGTGTCAAGTTGTGCCACAAAGTTATATCCATTTTAAATCCCATGTTATAGT 513
Db      16608 GCTTTACCAACCGTCTTGTGGCAGCTTTGCT---CAAGTATACGGATGATTCCTTG 16552
Qy      514 ACGAATTAAGATTTTGTGCAACAGTTCACAGATTCACATTTGTATATGCCAAAAAAG 573
Db      16551 CATTAACCCACCTTGTTCAACACCTGTCCAGGTGACACCTGCTGCGCACAGAAAG 16492
Qy      574 ATAGATCTGGGTTTGTATGTTGCAATTCGCAATTTATATGCTGATTTATATAGAAATT 633
Db      16491 ATTGGTCTGGATTGAGTGTGCTTCCGCTTTGTGCTCTTATGATGAGTTTC 16432
Qy      634 CAGCCAGCTTTGATTAATGACGTGTTCAGGTTCTAGAAAGTATCTGAGAAAGTTCCC 693
Db      16431 CCGCGGAGTCCGTAACATGTCATTGACAGTGAAGGA-----CCTCGAATACGGG 16378
Qy      694 ACAGAGTTGAAAAAATGATTGAAGTACGTGGGAATTCACATGAAGATGTACTTA 753
Db      16377 GCTGTGTTGAAACCTACCGTTATCAAAAGTGAAGTGTCTGGAACCATCTTCTTG 16318
Qy      754 CCAATACGAATCAAGTTATTAATGAGTGTGACGTCAAGGCTGTGAGAAACACCAATTG 813
Db      16317 CCGCGGGAATCAACGCTGTTATGAGAGAGTCAAGGAGATCTGAGATCCAGGTATG 16258
Qy      814 GTATCAGAGTACTCCAAATGAAAAAGAAAGCCAGAAAGAGCTGTGTGTATGAC 873
Db      16257 GTGCCCAAGTGTATGATGATGCGAAGCAAGCAAGCCCGAGAAACCGAATGTGTGAGA 16198
Qy      874 CAGTTAATATGTCGAATTTACAGTTATGAAGAAATTAAGGAAATGCGTGAATAAC 933
Db      16197 GATCTCAACGCTGCCAATCATGCTCATGTTTCAAGATTCCTGCGCAAGCTCTT 16138
Qy      934 GACTCAGACCCAGAACTTATATTAAGATTAGATCATCTGTGTGAGCTTGAAGCTTT 993
Db      16137 CTACCTAACAGAGGCTTACGAACAACTTTGGCCGAGAGCTGCTCTTCAACGCTTA 16078
Qy      994 GCGATTAAAGAAATCAAGAAAGGTTTACAAAGCTTAACAA 1035
Db      16077 AAGATGATATATTTGACGAACCTCTCGAGAACTTGACGA 16036

```

RESULT 16
SPAC343/c 42947 bp DNA linear PLN 20-JUN-2003
LOCUS
DEFINITION S.pombe chromosome I cosmid c343.
ACCESSION AL109739
VERSION AL109739.1 GI:5706501
KEYWORDS
SOURCE Schizosaccharomyces pombe (fission yeast)
ORGANISM Schizosaccharomyces pombe
Bukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomyces
1 (bases 1 to 42947)

REFERENCE
AUTHORS
Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R.,
Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D.,
Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T.,
Churcher, C., Collins, M., Connor, R., Cronin, A., Davis, P.,
Feltwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D.,
Hidalgue, J., Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S.,
Huckle, E.D., Hunt, S., Jags, K., James, K., Jones, L., Jones, M.,
Leather, S., McDonald, S., McLean, J., Mooney, P., Moulé, S.,
Munhall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neill, S.,
Pearson, D., Quail, M.A., Rabinowitch, E., Rutherford, K., Rutter, S.,
Saunders, D., Seeger, K., Sharp, S., Skelton, J., Simmonds, M.,

JOURNAL	TITLE	Squares, R., Squares, S., Stevens, K., Taylor, K., Taylor, R.G., Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J., Volckart, G., Aert, R., Robben, J., Grymomez, B., Weltjens, I., Vanstrele, E., Rieger, M., Schaffer, M., Miller-Auer, S., Gabel, C., Ruch, M., Dusterhoft, A., Fritze, C., Holzer, E., Moesti, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehman, H., Reinhardt, R., Pohl, T.M., Eger, P., Zimmermann, W., Medler, H., Wambut, R., Purnelle, B., Goffeau, A., Cadieu, B., Dreno, S., Gloux, S., Lelaure, V., Mottier, S., Gallbert, F., Aves, S.J., Xiang, Z., Hunt, C., Moore, K., Hurst, S.M., Lucas, M., Rochet, M., Gallardin, C., Tallada, V.A., Garzon, A., Rhode, G., Daga, R.R., Cruzado, L., Umeaz, J., Sanchez, M., del Rey, F., Benito, J., Dominguez, A., Revuelta, J.L., Moreno, S., Armstrong, J., Forburg, S.L., Cerutti, L., Lowe, T., McCombie, M.R., Paulsen, I., Potahkin, J., Shpakovski, G.V., Uesery, D., Barrell, B.G. and Nurse, P.
MEDLINE	21848401	The genome sequence of <i>Schizosaccharomyces pombe</i>
PUBMED	11859360	
REFERENCE	2	(bases 1 to 42947)
AUTHORS	Murphy, L., Harris, D., Wood, V., Rajandream, M.A. and Barrell, B.G.	
TITLE	Submitted (03-AUG-1997) European Schizosaccharomyces genome sequencing project, Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: pombe@sanger.ac.uk	
JOURNAL	Notes: Details of <i>S. pombe</i> sequencing at the Sanger Institute are available on the World Wide Web. (URL, http://www.genedb.org/genedb/pombe/index.jsp) (URL, http://www.sanger.ac.uk/projects/S_pombe/) CDS are numbered using the following system eg SPAC343.01c. SP (S. pombe), A (chromosome 1), CSH10 (cosmid name), .01 (first CDS), c (complementary strand). However, clones may have been reorientated since the original submission, therefore the complementary strand notation may be invalid for strand inference. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once.	
FEATURES	Location/Qualifiers	
source	1..42947	
gene	/organism="Schizosaccharomyces pombe"	
	/mol_type="genomic DNA"	
	/strain="972h"	
	/db_xref="taxon:4896"	
	/chromosome="I"	
	/map="IL"	
	/clone="cosmid C343"	
	/complement(join(61..1333,1400..1407))	
	/gene="SPAC343.01c"	
	/complement(join(61..1333,1400..1407))	
	/gene="SPAC343.01c"	
	/note="phosphomevalonate kinase (predicted); similar to <i>S. cerevisiae</i> ERG8"	
CDS	/codon_start=1	
	/protein_id="CAB52264.1"	
	/db_xref="GI:5706502"	
	/db_xref="GOA:O9UT88"	
	/db_xref="TrEMBL:O9UT88"	
	/translation="MKVTCASAGKVLGAGVIVLDPQSGVIGITAKGYASTLLD KCGIVKVSPOFINAEMLNIDWTSPIRVQIYNCELEKPNPFVQLAFYVINYF FSTGRPLCWDLOVTLQVNDNAVYHQPOLKPDQSYPKFNPLNCTLGGVHKGGLSSA AMITSLGLFLSLRLTIDTDGSKLIDSTKVIYHNLAAJHCSAOGKVSQGVGVA AATGSCYLRDPDLTLEQLVYDEQIKNINFSLEKIVSKMSDVPVLPATYVC LMDGVAGSSRPGKVKVQKQKPEKSKCPDILSRVLSIKNCPLSSSLDSEL OSQPSRIRLORITVEAKVDLEPKQRTIILNIDLEQLREVIGVPGAGGPAQCLIA INHTIENVITKWDGVPMDVSPADGLAVE"	
misc_feature	/complement(1334..1353)	
	/gene="SPAC343.01c"	
	/note="ctaactagctctggcag, splice branch and acceptor"	
	/complement(1394..1395)	
	/gene="SPAC343.01c"	
	/note="gtaagc, splice donor sequence"	
	2277..2825	
gene	/gene="SPAC343.02"	
	/note="mitochondrial ribosomal protein S19; similar to <i>S. cerevisiae</i> IMG1"	
	/codon_start=1	
	/protein_id="CAB52265.1"	
	/db_xref="GI:5706503"	
	/db_xref="GOA:O9UT87"	
	/db_xref="TrEMBL:O9UT87"	
	/translation="MFNAKHFNPGLGPGWLOKDGISLKNRPNPKPQPKHVDL SLFEKCSVYDEGSEKRFMRHSPNRVVRGAVILVSYKSPKSDVNRAGVILR LRHQPSSILRNWGVGVLEPLIPYQIKRIVLKENGLSRPRAYLSTYLRP RFRPEPVSLVRKTIENQHKP"	
	2484..2750	
misc_feature	/gene="SPAC343.02"	
	/note="match to PF01245 Ribosomal_L19, Ribosomal protein L19 Score 24.79"	
	join(3630..3648,3692..3814,4032..4174)	
	/gene="SPAC343.03"	
	/note="synonym: APC11"	
	join(3630..3648,3692..3814,4032..4174)	
	/gene="SPAC343.03"	
	/note="anaphase-promoting complex (APC) (PMID 12477395); involved in cyclin degradation (required); involved in metaphase-anaphase transition (required); zinc finger protein; zF-C3HC4 type (RING finger); ubiquitin ligase (E3) (predicted); conserved eukaryotic protein; similar to <i>S. cerevisiae</i> APC11"	
	/codon_start=1	
	/protein_id="CAB52266.1"	
	/db_xref="GI:5706504"	
	/db_xref="TrEMBL:O9UT86"	
	/translation="MKVILRYHAIANTWPTPKDDVSGICRVPEDCGCPQDTSRGN CPTWGRKHIIFHACTIONMLATSGSQGCMNDQTFVVAISTNKSRTQ"	
	3649..3654	
misc_feature	/gene="SPAC343.03"	
	/note="gtaagc, splice donor sequence"	
	3681..3691	
misc_feature	/gene="SPAC343.03"	
	/note="ctaactagctag, splice branch and acceptor"	
	3815..3820	
misc_feature	/gene="SPAC343.03"	
	/note="gtaaga, splice donor sequence"	
	4019..4031	
misc_feature	/gene="SPAC343.03"	
	/note="ctaactagctag, splice branch and acceptor"	
	/complement(join(4455..5753,5878..6028,6083..6156))	
	/gene="SPAC343.04c"	
	/complement(join(4455..5753,5878..6028,6083..6156))	
	/gene="SPAC343.04c"	
	/note="WD repeat protein"	
	/codon_start=1	
	/protein_id="CAB52267.1"	
	/db_xref="GI:5706505"	
	/db_xref="TrEMBL:O9UT85"	
	/translation="VALDEKFOLEVIHLLQFLNDYGYDESILKALEKETGVSEEDV KRIKQAVIQGDMITAEAFSTIMOLRDSKREKQAEFLQKOCLEABAGALCEAYVL ONFESTDNEKMERIVSITLSSNNKSNNEITKRGYCNTRDILNQLSBYSPILLP KRLEHLIQQAKDQVSSQVHVNANKNSFLSDYKADSELPTEKHYFHHSDEWQ IYSHNGVYLASASDKRTAIIFDVNLRVRLIGHIDTVAYIRMSPDRLHLSQSCD KSVILMDLPTKEKLDYHGGSVSCCCMLPGLSFTTSPOCHITRMNLNGSLYKWE DVNITVDMATDGGTLYVGEFOLINADKHAIGNFTSPRECIKISLOSQVTSCLS KDSRYALNLNHPHTFMDLEENLIVQRMGHKIGNPLIGSCFGSKDTPVLSGEDD KIRIWRSGKLATLSHYVCNVAIVNPDPYQFASAGDNTYRISNNDNPRQ"	
	complement(4482..4601)	
misc_feature	/gene="SPAC343.04c"	
	/note="match to PF00400 WD40, WD domain, G-beta repeat"	
	complement(5130..5228)	
misc_feature	/gene="SPAC343.04c"	
	/note="match to PF00400 WD40, WD domain, G-beta repeat"	
	complement(5256..5360)	
misc_feature	/gene="SPAC343.04c"	
	/note="match to PF00400 WD40, WD domain, G-beta repeat"	


```

misc_feature      complement (5754..5769)
                  /gene="SPAC343.04c"
                  /note="ctacatgctactag, splice branch and acceptor"
misc_feature      complement (5872..5877)
                  /gene="SPAC343.04c"
                  /note="gtacac, splice donor sequence"
misc_feature      complement (6029..6046)
                  /gene="SPAC343.04c"
                  /note="ctacatgctactgtag, splice branch and acceptor"
misc_feature      complement (6077..6082)
                  /gene="SPAC343.04c"
                  /note="gtacac, splice donor sequence"
gene              complement (6814..6884)
                  /gene="SPATRAGLY.01"
                  complement (6814..6884)
                  /gene="SPATRAGLY.01"
                  /product="tRNA-Gly"
                  /note="tRNA Gly anticodon GCC, Cove score 72.94"
                  /evidence=not experimental
                  join(7781..7783,7839..7938,7981..9737)
                  /gene="vma1"
                  /note="synonym: SPAC343.05"
                  join(7781..7783,7839..7938,7981..9737)
                  /gene="vma1"
                  /EC number="3.6.3.14"
                  /note="ty-type ATPase; H(+)-ATPase (PMID 1441756); vacuolar
                  ATP synthase (catalytic subunit A) (PMID 1441756); similar
                  to S. cerevisiae TFP1"
                  /codon_start=1
                  /protein_id="CAB52268.1"
                  /db_xref="GI:5706506"

Query Match      6.7%; Score 86.6; DB 8; Length 42947;
Best Local Similarity 56.5%; Pred. No. 5.1e-09;
Matches 205; Conservative 0; Mismatches 149; Indels 9; Gaps 2;

QY 499 CCCAATGTTATCACTAGCAATTAAGATATTTTGCACACGTTGCACAGATTGCACTTGT 558
DB 786 CTCAAAATTGATGATTAACCAAGTCATAGTTCAACATCTGCACAAAATGCCACTGT 727
QY 559 TATGCCCAAAAAAGATAGATCTGGTTGATGTTGCAACTGCAATTTATGCTGATT 618
DB 726 TCTGCACAGAAAAGTAGAAGTGAGTGAGTTGAGTGCGCTGCCACTGGGAAAGTTGT 667
QY 619 GTATATGAAGATTTCAGCCAGCTTGATTAATGACGTGT-----TCAGTTCTAGAA 672
DB 666 ATTTATAGAGGTTGATCTTAATTAATTGAACAACCTGCTGCCATPACGACAA 607
QY 673 AGTATCTCGAAGAGTTCCCAACAGGTTGAAAAAATTGATTGAAGTAACTGGAAATTC 732
DB 606 ATAAAGAACATTAATTTTCCACCGAATCGAAAAAATAGATTTTAAAAAATGG---TCA 550
QY 733 AAACATGAAGAATGTACATTACCATACGGAATCAATTATTAATGCGGTGAGTCAAGGT 792
DB 549 GATGTTGTTCCATTCAATTTGCTCTGCTACTTATTTTGTCTTAATGGGGAGTGCTGT 490
QY 793 GGCTCAGAAACCCCAATTTGTTATCGAGTACTCCAAATGAAAAAGAAAAAGCCAGAA 852
DB 489 GGCAGTAGACACCGGGTAGTGTAAAAAAGTAGACGATGCAAAAAGAAAAATCCAGAG 430
QY 853 GAA 855
DB 429 GAA 427

```

```

RESULT 17
AX485723/c
LOCUS
DEFINITION Sequence 3023 from Patent W002053728.
ACCESSION AX485723
VERSION AX485723.1 GI:22319939
KEYWORDS

```

```

SOURCE
ORGANISM Candida albicans
REFERENCE
AUTHORS Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
TITLE Gene disruption methodologies for drug target discovery
JOURNAL Biotra Pharmaceuticals, Inc. (US)
FEATURES
source
1..65
/organism="Candida albicans"
/mol_type="unassigned DNA"
/db_xref="taxon:5476"

```

```

ORIGIN
Query Match      5.0%; Score 65; DB 6; Length 65;
Best Local Similarity 100.0%; Pred. No. 0.00063;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTCAAAAGCATTTAGTGACCTGGAAGCAATTTCTTCTGCTGATATTGTTCTT 60
DB 65 ATGTCAAAAGCATTTAGTGACCTGGAAGCAATTTCTTCTGCTGATATTGTTCTT 6
QY 61 GAGCC 65
DB 5 GAGCC 1

```

```

RESULT 18
AX241110/c
LOCUS
DEFINITION Sequence 348 from Patent W00160975.
ACCESSION AX241110
VERSION AX241110.1 GI:15797985
KEYWORDS
SOURCE
ORGANISM synthetic construct
artificial sequences.
REFERENCE
1 Roemer,T., Jiang,B., Boone,C. and Bussey,H.
TITLE Gene disruption methodologies for drug target discovery
JOURNAL Patent: WO 0160975-A 348 23-AUG-2001,
Biotra Pharmaceuticals, Inc. (US)
FEATURES
source
1..90
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="DNA primer"

```

```

ORIGIN
Query Match      5.0%; Score 65; DB 6; Length 90;
Best Local Similarity 100.0%; Pred. No. 0.00061;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTCAAAAGCATTTAGTGACCTGGAAGCAATTTCTTCTGCTGATATTGTTCTT 60
DB 65 ATGTCAAAAGCATTTAGTGACCTGGAAGCAATTTCTTCTGCTGATATTGTTCTT 6
QY 61 GAGCC 65
DB 5 GAGCC 1

```

```

RESULT 19
CNS06KLC
LOCUS
DEFINITION T3 and of clone AT0AA006B09 of library AT0AA from strain CBS 4311
of Saccharomyces servazzii, sequence tagged site.
ACCESSION AL403030
VERSION AL403030.1 GI:12162807
KEYWORDS

```

SOURCE
ORGANISM Saccharomyces servazzii
Saccharomyces servazzii
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE
AUTHORS 1 (bases 1 to 963)
Soucier,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Maupertuy,A., Neuvéglise,C., Olier-Kalogeropoulos,O., Potier,S.,
Sautin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Winker,P. and Weissbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)

TITLE
JOURNAL 20584711
MEDLINE 1152876
PUBMED 20584711

REFERENCE
AUTHORS 2 (bases 1 to 963)
Casaregola,S., Lepingle,A., Bon,E., Neuvéglise,C., Nguyen,H.,
Artiguenave,F., Winker,P. and Galliardin,C.
Genomic exploration of the hemiascomycetous yeasts: 7.
Saccharomyces servazzii
FEBS Lett. 487 (1), 47-51 (2000)

TITLE
JOURNAL 20584717
MEDLINE 1152882
PUBMED 20584717

REFERENCE
AUTHORS 3 (bases 1 to 963)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,
2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
bege@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This STS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces marxianus, Kluyveromyces
lactis var. lactis, Kluyveromyces fragilis var. fragilis, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
source location/Qualifiers
1..963
/organism="Saccharomyces servazzii"
/mol_type="genomic DNA"
/strain="CBS 4311"
/db_xref="taxon:27293"
/clone="AT0AA006B09"
/clone_1fb="AT0AA"
/note="end : 73"
misc_feature <817..>954
/note="similar to Saccharomyces cerevisiae ORF YMR220w [
ERG8 : phosphomevalonate kinase]"
/evidence="not_experimental"

ORIGIN
Query Match 4.3%; Score 56; DB 11; Length 963;
Best Local Similarity 66.7%; Pred. No. 0.051; Indels 0; Gaps 0;
Matches 80; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
4 TCAAAAGCATTTAGTGCACCTGGAAGCAATTTCTTGCTGATATTGTTCTTGAG 63
DB 832 TCGAGAGCATTTAGTACCCCGGGAAGCCCTTATTATGATGATGCTGTTTGGAT 891
QY 64 CCAATTATGATGCTTATGATGACAGCATTTGATGACGAATGATGATTTATACACA 123
DB 892 CCACATTAACAGTCAATTCGTGTTAGCTTTTCAGCGAAGATGATGCTGTGTAAGACA 951

RESULT 20
CNS06J37 1037 bp DNA linear STS 30-NOV-2001
LOCUS T7 end of clone AS0AA025E12 of library AS0AA from strain CLIB 533
DEFINITION of Saccharomyces bayanus, sequence tagged site.

ACCESSION AL401081
VERSION AL401081.1 GI:12158232
KEYWORDS STS.
SOURCE Saccharomyces bayanus
ORGANISM Saccharomyces bayanus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE
AUTHORS 1 (bases 1 to 1037)
Soucier,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Maupertuy,A., Neuvéglise,C., Olier-Kalogeropoulos,O., Potier,S.,
Sautin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Winker,P. and Weissbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)

TITLE
JOURNAL 20584711
MEDLINE 1152876
PUBMED 20584711

REFERENCE
AUTHORS 2 (bases 1 to 1037)
Bon,E., Neuvéglise,C., Casaregola,S., Artiguenave,F., Winker,P.,
Aigle,M. and Durrens,P.
Genomic exploration of the hemiascomycetous yeasts: 5.
Saccharomyces bayanus var. uvarum
FEBS Lett. 487 (1), 37-41 (2000)

TITLE
JOURNAL 20584715
MEDLINE 1152880
PUBMED 20584715

REFERENCE
AUTHORS 3 (bases 1 to 1037)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,
2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
bege@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This STS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces marxianus, Kluyveromyces
lactis var. lactis, Kluyveromyces fragilis var. fragilis, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
source location/Qualifiers
1..1037
/organism="Saccharomyces bayanus"
/mol_type="genomic DNA"
/strain="CLIB 533"
/variety="uvarum"
/db_xref="taxon:4931"
/clone="AS0AA025E12"
/clone_1fb="AS0AA"
/note="end : 77"
misc_feature <3..>560
/note="similar to Saccharomyces cerevisiae ORF YMR220w [
ERG8 : phosphomevalonate kinase]"
/evidence="not_experimental"

ORIGIN
Query Match 4.3%; Score 55.6; DB 11; Length 1037;
Best Local Similarity 57.5%; Pred. No. 0.063; Indels 0; Gaps 0;
Matches 100; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
986 TGACCTGTCGATTAAGAACATCAGAAAGGCTTACAAGCATTAACAAAATCAGAGC 1045
DB 256 TGAAGAGATTCGTGTGACAGATTAAGCGTTGTTTACAAAATTAATTAAGATCTGGAG 315
QY 1046 TTCCAAATTGAACCTGATGTCGAACCCAGATTGTTGAGCGTTGCAAGATTCGTGTT 1105

AUTHORS	Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Anka, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Otsuka, R., Nakazawa, H., Takamiya, M., Kato, Y., Yoshizawa, T., Tanaka, T., Kudo, Y., Yamazaki, Y., Nishida, N., Oguchi, A., Aoki, K., Masuda, S., Yamagi, M., Nishimura, M., Yamagishi, A., Oshima, T. and Kikuchi, H.
TITLE	Complete genome sequence of an aerobic thermocacidophilic crenarchaeon, <i>Sulfolobus tokodaii</i> strain7
JOURNAL	DNA Res. 8 (4), 123-140 (2001)
MEDLINE	21456156
PIUMED	11572479
REFERENCE	2 (bases 1 to 253050)
AUTHORS	Director-General of Biotechnology Center, Kawarabayasi, Y., Tanaka, T., Hino, Y. and Kikuchi, H.
TITLE	Direct Submission
JOURNAL	Submitted (22-DEC-1999) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan (/e-mail:bioinfo.go.jp, Tel:81-3-3481-1933, Fax:81-3-3481-8424) Kawarabayasi, Y. is officially affiliated with the National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki 305-0065, Japan
COMMENT	Yamagishi, A. and Oshima, T. are at Tokyo University of Pharmacy and Life Science, Hachioji, Tokyo 192-0392, Japan The other authors are at the National Institute of Technology and Evaluation, Nishihara 2-49-10, Shibuya-ku, Tokyo 151-0066, Japan Genbank release 109; EMBL release 56.0; SwissProt release 38.0; PIR-Protein release 62.0; and OWL release 31.4. E-mail address for comments and questions: kyutaka@ite.go.jp ORF Organization, sequence alignment and more information are available at W.W.W. site of Biotechnology Center, URL: http://www.bio.nite.go.jp/dogan/genome_list-e.html/ .
FEATURES	location/Qualifiers
source	1..253050
	/organism="Sulfolobus tokodaii str. 7"
	/mol_type="genomic DNA"
	/strain="7"
	/db_xref="taxon:273063"
gene	480..1019
	/gene="ST0846"
CDS	480..1019
	/gene="ST0846"
	/codon_start=1
	/transl_table=1
	/product="179aa long hypothetical protein"
	/protein_id="BAB5859.1"
	/db_xref="GI:15621866"
	/translation="MKDDIYKVEDYVEARVIESDILLSTLMKEGYTRNSAGKAI NIVKVALMSALVYTNDEKLALADCKDEREMIKKATVPTSHVYLAQMLKDVGIDIDI MAYKALDEKIDNRGNVTL"
gene	complement (1500..2816)
	/gene="STyep24"
	complement (1500..2816)
repeat_region	complement (1500..2816)
	/note="dispersed repetitive unit, subtype IT"
gene	/rpt_type="dispersed"
	/rpt_unit="1500..2816"
CDS	1621..2613
	/gene="ST0847"
	1621..2613
	/gene="ST0847"
	/note="PIR:S74012 percent identity: 33.935 in 277aa."
	/codon_start=1
	/transl_table=1
	/product="330aa long conserved hypothetical protein"
	/protein_id="BAB65860.1"
	/db_xref="GI:15621867"
	/translation="MGIEIYOSVLPQKLYGWSKYLDCCEILKEILDVIRREGRK /FOYDPLTVIKMLIVMLYRISYRGDTILATNIIVVRPIKVKIIPKSSILNMFVVKF SRIEDLLKVRFPRESLVEERQLPTHLIAEFGGDIRLDSFVPELNGKKSJET HREKLLDLPRTDSKRPKHVTLGI DEKQKRR"

gene
3365..3868
/gene="ST0848"
3365..3868
/gene="ST0848"
/note="PIR:S74037 percent identity: 43.293 in 164aa."
/codon_start=1
/transl_table=11
/product="167aa long conserved hypothetical protein"
/protein_id="BAB65861.1"
/db_xref="GI:15621868"
/translation="MDTSFLIDWVKYDKRDLFDYTTVPFSESVANEIRRTPLMT
SEWLAKRIKILEETVDARKALVDDMTREPLSADYPAVCLVLGKELNDVLT
NGVFAKEIIEBNSVNYRGIDILYIISOKGLSDFINVKYKINSTKHYSKRIR
EKYGIEI"
4482..5090
/gene="ST0849"
4482..5090
/gene="ST0849"
/note="PIR:T00006 percent identity: 53.247 in 77aa."
/codon_start=1
/transl_table=11
/product="202aa long hypothetical pepk protein"
/protein_id="BAB65862.1"
/db_xref="GI:15621869"
/translation="MGITKDISKNDIKNAIIEIYSGKSTPSKOBELLEKKRSNTE
IDNIIEDLVKGFIIKKIKGKNKKEIOMFIEGKVDENIGEDTIEGSDQIKAL
ESAPIVLSKYERLIEDLKITIPDRIPTEPLDRVYEFANSLGHASIKDLRIQLG
LSBERFMRYFRYIIIONYELIIGGEGFIRKGVMTGIIRKX"
5099..6433
/gene="ST0850"
5099..6433
/note="motif=ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="444aa long hypothetical protein"
/protein_id="BAB65863.1"
/db_xref="GI:15621870"
/translation="MCNPSYVARVPDNNVILNGVMSARKSTIANIINDGWT
ITGSPGKTTITLANVVKDKRDVFIYLLSDRISAKAMYTIRKLEYVSKV
FNYSLHKKEIGYKGFYISKDFKNMLHQESSDDARIKLYCRYKEDIGLEFL
NDLKTQSLFKVGLIDETSKKEIIDEIHELINKYKIPATILIDEVKVNALVR
RLENOIIEIKDKDEIKLKAECERYKELIDERNALNLDARPKYSEAL
NECKNGAYIYDECIKDYLKPYIDESLDDYSTALEKRIKLEIKQAPVYKI
HBRCKLDEYCTVDIIPETENTYIYGDKVSNBENKRIKVKGLUENQIYKDKKV
KRTVNNVNSDPTIIPVNNLSIKTIYDNDKERDLVPMNLRLGIYI"
6481..6894
/gene="ST0851"
6481..6894
/gene="ST0851"
/codon_start=1
/transl_table=11
/product="137aa long hypothetical protein"
/protein_id="BAB65864.1"
/db_xref="GI:15621871"
/translation="WYMISRGIVILTFPFVALLLELMIDPSQIKTSPGMKEF
TPTVSLYFRVYTYLFRAPSYLASKKLPSTWETIYLTIGSTLEDDIYIWLVRVPS
WAMFYVYIGIDIVGLMLIRKIKKEKVR"
complement(7523..8800)
/gene="ST0852"
complement(7523..8800)
/note="Swiss_Prot:Q10858 percent identity: 28.289 in
304aa."
/codon_start=1
/transl_table=11
/product="425aa long hypothetical transporter"
/protein_id="BAB65865.1"
/db_xref="GI:15621872"
/translation="WBEKTLSSQALALGNIIGAGITVMAGVSTTAAGPALAFL
ITAVYAMSGLNNELASVFPVEGGVYSFALLSGETIGFVGMFRVIGVATISGAT
ALGFSGLITTPSLPSFLYLLAILIIVLIIIDYIGLTLAAIEISITVVIINILGVI

gene
8837..9703
/gene="ST0853"
8837..9703
/gene="ST0853"
/note="PIR:C75043 percent identity: 28.821 in 229aa."
/codon_start=1
/transl_table=11
/product="288aa long hypothetical protein"
/protein_id="BAB65866.1"
/db_xref="GI:15621873"
/translation="MILGKLAELKRGYKIYGYEKLSTONIAVKMLVKNYGLK
RISKRIGEISLKYKGDKEVDKPTGEGEPIKLPKYPILVIDMSLPNEDE
BEKKTUQVNTLVIRKFLMDGALINSNISLGRARNIDNETDNCIYLDPRGD
IIANEIIRTDVFIIGGIVDKGRILKATSLAKRYPKCKVITLRGSIIVPDEIN
KIADIILGVKPKGELEKKAIIETQSNDSKIRILVDVNORGLIEEYVMANOKY
KLIRLIGVNAS"
complement(9806..10966)
/gene="ST0854"
complement(9806..10966)
/gene="ST0854"
/note="Swiss_Prot:P19780 percent identity: 35.897 in
117aa."
/codon_start=1
/transl_table=11
/product="386aa long hypothetical transposase"
/protein_id="BAB65867.1"
/db_xref="GI:15621874"
/translation="MTKDFVSKAFADISQSLTAKGELVIKOEKPSVTVKEVEFP
YNNIEILIKLEEGVNGIIEATGVFPYILKLEGGFKTVVNPALTEILGKT
DKLDQRLVAVMTGVIVGYSYIPGEIKELRETPHRELVNKTIOVREIKTEIA
GKTIKPFDKGRQLEKLAKGEBLSKESEBELKELGRLNDAEKLTKQVLELKS
ENMKEVDMISIKIPKVIETLSKIPGILTAATVIAEPGDISRPNSKARAYAGF
APRTQSGNSHSGMIRGNKRLRGFPYLAASARMLPEFNEYSRLIARGSVQAT
CALAGKACIVYHVLKGVYKGIYVKRLRIPKGEVNVDFVGDALDSLSP"
10218..10748
/gene="ST0855"
10218..10748
/gene="ST0855"
/note="PIR:F71169 percent identity: 37.313 in 67aa."

Query Match 4.0%; Score 51.8; DB 1; Length 253050;
Best Local Similarity 64.7%; Pred. No. 0.27; Mismatches 42; Indels 0; Gaps 0;
Matches 77; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 517 AATAAGATATTTTGCACACGTTGCACAGATTTGATGCCCAAAAAAGATA 576
DB 134901 AACAAAGATGAATCTATTAATATAGCACGAAGAAATTAATAGCAAAAGATATT 134842

QY 577 GGATCTGGCTTGATGTTGCAACGCAATTATATGCTCGATTTGATATGAAGATTCA 635
DB 124841 GGAAGCGTTTGGATATAGCCTCAGCTGTTATGTTCAATAGATTATGAAGATTTA 124783

RESULT 23
AC009757/c AC009757 166918 bp DNA linear PRI 27-AUG-2002
LOCUS Homo sapiens chromosome 5, clone RP11-37w16, complete sequence.
DEFINITION AC009757
ACCESSION AC009757
VERSION AC009757.8 GI:8072443
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 166918)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 5, clone RP11-37w16
JOURNAL Unpublished

REFERENCE
AUTHORS

2 (bases 1 to 166918)

Birtten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benu,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,B., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heatford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrum,J., Molla,M., Morris,W., Morrow,N., Mychalackyj,J.,
Naylor,T., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Roney,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tessfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (31-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 166918)

Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,N.,
Murphy,T., Naylor,T., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (09-JUN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 166918)

Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,N.,
Murphy,T., Naylor,T., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
5 (bases 1 to 166918)

Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,

TITLE
JOURNAL
COMMENT

Submitted (27-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced g1:7960347.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

FEATURES
source

Location/Qualifiers
1..166918
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/map="5"
/clone="RP11-37M16"
/clone_1lb="RPCT-11 Human Male BAC"
complement(12..433)
/rpt_family="L1P2a"
583..891
/rpt_family="AlusX"
1655..1819
/rpt_family="L2"
1834..1854
/rpt_family="AT_rich"
3154..3221
/rpt_family="AT_rich"
4414..4472
/rpt_family="AT_rich"
5561..7641
/rpt_family="AT_rich"
7642..7662
/rpt_family="AT_rich"
7707..7763
/rpt_family="AT_rich"
7764..8129
/rpt_family="Tigert1"
8134..8519
/rpt_family="Tigert1"
8520..8810
/rpt_family="AlusG"
8811..9500
/rpt_family="Tigert1"
9515..10051
/rpt_family="L1aP"
10111..10149
/rpt_family="AT_rich"


```

Db      291 AGAGAGACATGATCATGACAAAAGAGACAGCCAAAAGCTTTAGAGACAAATTAC 350
Qy      969 TCATTCTGTGAGCCTTTGACCTGTGCGATTAAAGAACAGAAAAGGTTCAAGCATT 1028
Db      351 GGATTAAGCAAGAACATCTTACGCGTCCCTTGCCTATATCAATGAAAAAGACAGAGC 410
Qy      1029 AACACAAAATCAGAGGTTCCATTGAACCTGATGTCACAAACCAAGTTGTGACCGTTG 1088
Db      411 AAAAGAAAAAGAGCAGACAGAAAAAATTAAAGAAAGAAAAACAATCTCAGAGCGAG 470
Qy      1089 TCAAGAGATTCCTGTTGTGTTGGTGGTGTGTTCCAGGTGCTGTGATACAGTCAAT 1148
Db      471 CCGTCAAGGCTCTTCCCGCTGACTTGACCATCACTGAGAGCTAAGAAACAAGTTGAAA 530
Qy      1149 AGCTGATTAAGTGTGGAATAATCAAGTGGAAATTTAAGCAGAAA 1195
Db      531 ACATTAGCAAACTTGACTGCTGAACTGATTAAGTTAAAGAGAAA 577

RESULT 27
SPERM41      1288 bp      DNA      linear      BCT 12-MAY-1993
DEFINITION   S.pyogenes for emm41 gene.
ACCESSION    X58178
VERSION       X58178.1 GI:47362
KEYWORDS      emm41 gene; M protein; PCR; virulence factor.
SOURCE        Streptococcus pyogenes
ORGANISM      Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
               Streptococcus.
REFERENCE     1 (bases 1 to 1288)
AUTHORS       Podbielski, A.
TITLE          Three different types of organization of the vir regulon in group A
               streptococci
JOURNAL       Mol. Gen. Genet. 237 (1-2), 287-300 (1993)
MEDLINE       93204905
PUBMED        8455563
REFERENCE     2 (bases 1 to 1288)
AUTHORS       Podbielski, A.
TITLE          Direct Submission
JOURNAL       Submitted (28-FEB-1991) A. Podbielski, Institut f Med
               Mikrobiologie, an der Rhein Westf Techn Hochschule, Pauwelsstr
               Klinikum, 5100 Aachen
COMMENT        See also X58157-X58168 for related sequences.
FEATURES      location/Qualifiers
               1..1288
               /organism="Streptococcus pyogenes"
               /mol_type="genomic DNA"
               /strain="C101/103/2"
               /db_xref="taxon:1314"
               17..1256
               /gene="emm41"
               17..23
               /gene="emm41"
               30..1256
               /gene="emm41"
               /codon_start=1
               /transl_table=11
               /product="M protein type 41"
               /protein_id="CAA41167.1"
               /db_xref="GI:47363"
               /db_xref="GOA:O54837"
               /db_xref="TrEMBL:O54837"
               /translation="MIRKDTNRHSLRKIKTGTSAVAAVAALVLAGFANOTVEKAGN
               ALAQAEALADVLNNTPHNQLRDPVGAFFRNNELEKI IQEKRELEGGLDAELKR
               LNEBHHDHDKAEARKKALEDKLADQEHLDGALRYINKEAEKKEKEAEOQKLKEEKQ
               ISDASROGLRDLDSAEAKQVEKDLANLTAPDPKREKOVSDASROGLRDLSDAS
               BEAKQVEKDLANLTAPDPKREKOVSDASROGLRDLSDASBEAKQVEKALREERAS
               KLAAEKLNKLEBESKULTPEKSKALQALCEBAKALQEQALQAEELAKLPAEKAQSG
               SOTPAKFGKTAVPKGGAPQAGTKPNQKAPMKETKQQLPSTGSAANPFTYAAALTV
               MATAGVAAVVKKKEEN"
               repeat_region
               507..827

```

```

misc_feature      /note="C repeat"
                   576..632
                   /gene="emm41"
                   /note="spacer 1"
misc_feature      702..758
                   /gene="emm41"
                   /note="spacer 2"

ORIGIN
Query Match      3.7%; Score 47.8; DB 1; Length 1288;
Best Local Similarity 46.5%; Pred.No.3.4; Indels 0; Gaps 0;
Matches 154; Conservative 0; Mismatches 177;

865 GTGTATGACCAAGCTTAATAGTCCCAATTACAGTTATGAGAAATGCGT 924
Db      300 GAGTTAGAAAGCGCTTAAGATGCTGAGTTGAAGCAGCTTAAGAGAGAGATGATCAT 359
Qy      925 GAAAAATACGACTCAGACCCAGAGACTTATATTAAGAAGTTGATCATTTGTTAGCCT 984
Db      360 GACAAAAAAGAGCAGAGCAAAAGCTTTAGAGACAAATTAGCGATTAACAGAACAT 419
Qy      985 TTGACTGTTCGATTAAGAACATCAAGAAAGGCTTAACAGCTTAACCAAAATCAGAG 1044
Db      420 CTGACGGTGCCCTTCGCTATATCAATGAAAAAGAGCAGACGAAAGAAAAAGAGCA 479
Qy      1045 GTTCCAAATGAACCTGATGTCCAAACCAAGTTGTGACCGTTGTCAAGAGATTCTCGT 1104
Db      480 GAGCAGAAAAAACTTAAGAGAAAAAACAATCTCAGACGCAAGCCGTCAAGCTCTTCT 539
Qy      1105 TGTGTTGTGTGTGTGTTCCAGGTGCTGTGTGATACATGCAATAGCTGTATTAGTTG 1164
Db      540 CGTGACTTGAGCGCATCACTGAGACCTTAAGAAACAGTTGAAAAAGATCTACCAACTTG 599
Qy      1165 GAAATCAAGTCGGAATTTTAAGCAGAAA 1195
Db      600 ACTGCTGAATTTGATTAAGATTAGAGAGAAA 630

RESULT 28
AY139408      949 bp      DNA      linear      BCT 14-NOV-2002
DEFINITION   Streptococcus pyogenes isolate NS2660 M protein (emm) gene, emm70
               type, partial cds.
ACCESSION    AY139408
VERSION       AY139408.1 GI:24954582
KEYWORDS      Streptococcus pyogenes
SOURCE        Streptococcus pyogenes
ORGANISM      Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
               Streptococcus.
REFERENCE     1 (bases 1 to 949)
AUTHORS       Dyall-Smith, M.L., Krejany, S., Sriprakash, K.S., Delvecchio, A. and
               McMillan, D.
TITLE          emm sequences of Australian GAS isolates
JOURNAL       Unpublished
AUTHORS       Dyall-Smith, M.L., Krejany, S., Sriprakash, K.S., Delvecchio, A. and
               McMillan, D.
TITLE          Direct Submission
JOURNAL       Submitted (06-AUG-2002) Microbiology and Immunology, University of
               Melbourne, Grattan Street, Parkville, Victoria 3010, Australia
               Sequence identified in 2002 by students participating in the
               Molecular Microbiology Techniques Course (526-321), Department of
               Microbiology and Immunology, University of Melbourne, Australia
FEATURES      location/Qualifiers
               1..949
               /organism="Streptococcus pyogenes"
               /mol_type="genomic DNA"
               /isolate="NS2660"
               /db_xref="taxon:1314"
               /country="Australia: Northern Territory"
               <1..949
               /gene="emm"

```


SOURCE
 Mus musculus (house mouse)

ORGANISM
 Mus musculus
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 (bases 1 to 220761)
 Birren, B., Nusbaum, C. and Lander, E.
 Mus musculus chromosome 14, clone RP23-246A2

JOURNAL
 unpublished
 2 (bases 1 to 220761)

TITLE
 Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
 Anderson, S., Barina, N., Bastien, V., Boguslavsky, L., Boughalter, B.,
 Brown, A., Camarata, J., Campiolo, A., Chang, J., Chararo, B.,
 Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., Daretliano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gird, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kamat, A., Karatas, A., Kells, C., Labocque, K.,
 Lamasare, R., Landers, T., Labocque, J., Levine, R., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McSwan, P., McKernan, K., McPheters, R., Meldrum, J.,
 Menais, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Riback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Scjapanovic, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strassan, N., Subramanian, A., Talamas, J., Testaye, S., Theodores, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
 Direct Submission

JOURNAL
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE
 3 (bases 1 to 220761)
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
 Anderson, M., Arachchi, H.M., Barina, N., Bastien, V., Bloom, T.,
 Boguslavsky, L., Boughalter, B., Camarata, J., Chang, J., Choepel, Y.,
 Collymore, A., Cook, A., Cooke, P., Corum, B., Daretliano, K.,
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
 Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
 Graham, L., Grand-Pierre, N., Hafetz, N., Hagopian, D., Hagos, B.,
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
 Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
 Meldrum, J., Menais, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Schauer, S., Schauer, R., Seaman, S., Severy, P., Smith, C.,
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
 Talamas, J., Testaye, S., Theodores, J., Topham, K., Travers, M.,
 Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
 Direct Submission

JOURNAL
 Submitted (22-MAR-2004) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT
 On Mar 22, 2004 this sequence version replaced gi:2891236.
 All repeats were identified using RepeatMasker:
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

--- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 --- Project Information
 Center project name: L17243
 Center clone name: 246_A2

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces


```
Assembly program: Phrap; version 0.990339
Consensus quality: 210651 bases at least Q40
Consensus quality: 212745 bases at least Q30
Consensus quality: 214097 bases at least Q20
Estimated insert size: 225125; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
```

NOTE: Estimated insert size may differ from sequence length (see http://www.bsgc.bcm.tmc.edu/docs/genbank_draft_data.html)

NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
* 1 118527: contig of 118527 bp in length
* 118528 118627: gap of unknown length
* 118628 233719: contig of 115092 bp in length
Location/Contig/Size
```

FEATURES	Location/Qualifiers
SOURCE	1. .233719

```
misc_feature    1. .1277
```

```
misc_feature      1328.2414
```

misc_feature	6171.7063
--------------	-----------

```
/note="clone_boundary
clone_end:sp6
site:
end_sequence:BH287188"
```

ORIGIN

Query Match	3.5%	Score 45.8;	DB 2;	Length 233719;
-------------	------	-------------	-------	----------------

Matches 131; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

166 CCCCAATTGCAACGAGATGGGAATATCATATCATCAATACAGAGAAGCCGAGA 225

DB 172462 CCCACATTCGAAAGTGAGACTTTAATAATTGGGATACATTTTACAAGCAGATTGA 172521

220 GHAAGI I CAG I CACGCA I AHA I CCA I I I I AAGAG CCA I A I A I CCA I CG I I I AAG I I A I 285

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

.....

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

406 TCTCGTCCATTACCGAAGTGGAAGAACCGGA 438

Db 172702 GAGCCTCTCCTAACTGTATACAGAATGATCAGA 172734

AC125664/c

DEFINITION Rattus norvegicus clone CH230-126J2, WORKING DRAFT SEQUENCE, 4
unrounded bases

ACCESSION	AC125664
VERSION	3
GT	3052560

SOURCE Rattus norvegicus (Norway rat)
M10/ M105_FEMUR/ M105_PATE1/ M105_FEDIOF.

ORGANISM

ORGANISM	REFERENCE
<i>Rattus norvegicus</i>	Mizny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;	
<i>Rattus</i> .	
1. (bases 1 to 243143)	

AUTHORS

1 (bases 1 to 243143)

Muzny, D., Maer, Metzker, M., Lee, Abramzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Altshrook, S., Amin, A., Anguiano, D., Anyslebsch, V., Aoyagi, A., Ayodji, M., Bacé, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benhammed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dedetich, D., Delgado, O., Denison, S., Detanoo, C., Ding, Y., Dihn, H., Diya, K., Draper, H., Dugan-Socha, S., Dunn, A., Dudhik, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gablitz, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebrategeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, N., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levay, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Joranshubewa, L., Loulesged, H., Lozand, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawlinsey, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minga, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelamem, O., Okwunonu, G., Olarunpungson, A., Pal, S., Parke, K., Pasterkamp, S., Paul, H., Perez, A., Perez, L., Planckoch, C., Plapper, F., Polndexter, A., Popovic, D., Primus, E., Pu, L., L., Puauc, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojias, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smaj, S., D., Sneed, A., Sodergren, E., Sutton, X. Z., Sorelle, R., Sosa, J., Steagle, M., Strong, R., Sultun, A., Swakek, A., Taboc, P., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uমান, K., Vals, R., Vera, V., Villaana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, K., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, Y., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.

TITLES

REFERENCE

TITTLE

JOURNAL

0000000000

AUTHORS

**THE
JOURNAL**

COMMENT

[illegible]

Buck, J., Bowie, S., Brievara, M., Brown, E., Brown, M., Bryant, N. P.,
Buhay, C., Burch, P., Burkett, C., Burnett, K. L., Byrd, N. C.,
Carron, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhury, I., Christopoulos, C.,
Cleveland, C. D., Cox, C., Coyle, M. D., Dathorne, S. R., David, R.,
Davilla, M. L., Davis, C., Davy-Carroll, L., Dedertich, D. A.,
Delaney, K. R., Delgado, O., Demu, A. L., Ding, Y., Dinh, H. H.,
Douthwaite, K. J., Draper, H., Dugan-Rocha, S., Durbin, K. J.,
Earnhart, C., Edgar, D., Edwards, C. C., Elhaj, C., Emerling, S.,
Escotto, M., Fails, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P.,
Frantz, P., Gablisl, A., Gao, J., Garcia, A., Garner, T., Garza, N.,
Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Haylak, P.,
Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoggus, M.,
Holloway, C., Hollins, B., Homel, F., Howard, S., Huber, J., Huliyk, S.,
Hume, J., Ioshihara, I., Jackson, L. E., Jacobson, B., Jia, Y.,
Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U.,
Johnson, L., Kovach, J., Kovar, C., Katovic, J., Kureishi, A., Landry, N.,
Leal, B., Lee, E., Lewis, L. C., Lewis, L. J., Li, J., Li, Z., Licharge, O.,
Lieu, C., Liu, J., Liu, W., Louisse, H., Lozada, R. J., Lu, X.,
Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P.,
Merondel, I., Martin, R., Martindale, A., Martinez, E., Massey, E.,
McWhiney, E., McLeod, M. P., Meador, M., Mei, G., Metzger, S.,
Metzger, E., Miller, A., Miner, G., Miner, Z., Mitchell, T.,
Mondabac, K., Montgomery, K. T., Morgan, M., Morris, S., Moser, M.,
Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokwelu, S., Oguth, M., Okwunodu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojubakan, I., Rolfe, M., Ruiz, S., Saverly, G.,
Scherrer, S., Scott, G., Shen, H., Shim, C., Shoohtari, N., Sisson, I.,
Sodergren, E., Sonakke, T., Sparks, A., Stanley, H., Stone, K.,
Sutton, A., Svatek, A., Tabori, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Umanai, K., Vasquez, L., Vera, V., Villalton, D., Vinson, R., Wang, O.,

Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Kuchera-Lapate, R., Weinstein, G., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 105989)
Worley, K.C.
Direct Submission
Submitted (24-JUL-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 105989)
Worley, K.C.
Direct Submission
Submitted (27-FEB-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 105989)
Worley, K.C.
Direct Submission
Submitted (01-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Mar 1, 2003 this sequence version replaced gi:28570287.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Gene and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:

<http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES

source

Location/Qualifiers

1..105989

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="12"

/clone="RP11-407N8"

1..2004

/note="overlaps bases 194739..196742 of clone AC087884"

/function="clone overlap"

2205..2312

/rpt_family="MER99"

repeat_region 3552..3836
/rpt_family="AluX"
repeat_region 3837..3859
/rpt_family="TAAAA)n"
repeat_region 3883..3903
/rpt_family="AT_rich"
4027..4118
/standard_name="STS-241143"

STS

repeat_region complement(5005..5313)
/rpt_family="AluSg"
repeat_region complement(5410..5718)
/rpt_family="AluSx"
repeat_region 5756..5889
/rpt_family="AluJo"
repeat_region 5890..6177
/rpt_family="AluSx"
repeat_region 6178..6325
/rpt_family="AluJo"
repeat_region 6385..6584
/rpt_family="MLT21"
repeat_region 6585..6627
/rpt_family="(TCC)n"
complement(6628..6888)
/rpt_family="AluJo"
6669..6770
/standard_name="D11S2921"

STS

repeat_region 6900..6959
/rpt_family="MLT21"
repeat_region 7157..7282
/rpt_family="AluJo/FLAM"
7285..7592
/rpt_family="AluJo"
complement(7594..7860)
/rpt_family="AluSx"
8059..8129
/rpt_family="HERVL"
8130..8353
/rpt_family="MER20"
8354..9308
/rpt_family="HERVL"
9309..9619
/rpt_family="AluJo"
9620..9709
/rpt_family="HERVL"
9710..9827
/rpt_family="FLAM_C"
9838..10226
/rpt_family="HERVL"
10251..10871
/rpt_family="HERVL"
10873..11332
/rpt_family="MLT21"
complement(11333..11643)
/rpt_family="AluSx"
11644..11804
/rpt_family="MLT21"
11843..12017
/rpt_family="L2"
complement(12018..12313)
/rpt_family="AluSx"
12314..12473
/rpt_family="L2"
complement(12314..12473)
/rpt_family="AluSg"
12474..12498
/rpt_family="(TG)n"
12499..12794
/rpt_family="L2"
complement(12499..12627)
/rpt_family="AluSg"
13789..14044
/rpt_family="AluJo"
14782..15087

repeat_region /rpt_family="AluSx"
15312..15611
/rpt_family="AluSp"
repeat_region complement(15692..15996)

Query Match 3.5%; Score 45.4; DB 9; Length 105989;

Best Local Similarity 48.0%; Pred. No. 8;

Matches 130; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

624 TGAAGATTTCAGCCAGCTTTGATTAATGACGTTTCAGGTTCTAGAAAGTACCTGA 683

92156 TAAATATGACACAGTACTTTTAAATGTTATGCTGAATTTTATGTTAGACAAA 92097

684 GAAGTCCCAAGAGTTGAAAAATTTGATTGAAAGTAACTGGCAATTCAAAATGAAG 743

92096 CGAATTAATTCCTTTGTAATTAATCAACCAACTGAATTAATTAATGACCTAATGAAA 92037

744 ATGTACATTACATACGAAATCAAGTTAATGCGTACGTCAGCGGTGCTCAGAAAC 803

92036 TGTGTAATCACTAAGTCAAGTCACTCAATGAAAAAGGAAAGCCAGAAAGTCTCTG 863

804 ACCCAATTTGTTATCAGTACTCCTCAATGAAAAAGGAAAGCCAGAAAGTCTCTG 863

91976 TAAAGGAATGATCGTATGATTTCTTTTCAAAAAATTAACCAAGCTCAGAACAA 91917

864 TGTGTATGACACAGCTTAATATGCTCAATTTA 894

91916 TGAAGTTCTACATTAATACAGAGACAGATCA 91886

RESULT 37
AC009727_2
WPCOMMENT

Sequence split into 4 fragments LOCUS AC009727 Accession AC009727

Fragment Name Begin End

AC009727_0 1 110000

AC009727_1 100001 210000

AC009727_2 200001 310000

AC009727_3 300001 374634

Continuation 73 of 4) of AC009727 from base 200001 (AC009727 Homo sapiens chromosome 12

Query Match 3.5%; Score 45.4; DB 2; Length 110000;

Best Local Similarity 48.0%; Pred. No. 8;

Matches 130; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

624 TGAAGATTTCAGCCAGCTTTGATTAATGACGTTTCAGGTTCTAGAAAGTACCTGA 683

70307 TAAATATGACACAGTACTTTTAAATGTTATGCTGAATTTTATGTTAGACAAA 70366

684 GAAGTCCCAAGAGTTGAAAAATTTGATTGAAAGTAACTGGCAATTCAAAATGAAG 743

70367 CGAATTAATTCCTTTGTAATTAATCAACCAACTGAATTAATTAATGACCTAATGAAA 70426

744 ATGTACATTACATACGAAATCAAGTTAATGCGTACGTCAGCGGTGCTCAGAAAC 803

70427 TGTGTAATCACTAAGTCAAGTCACTCAATGAAAAAGGAAAGCCAGAAAGTCTCTG 863

804 ACCCAATTTGTTATCAGTACTCCTCAATGAAAAAGGAAAGCCAGAAAGTCTCTG 863

70487 TAAAGGAATGATCGTATGATTTCTTTTCAAAAAATTAACCAAGCTCAGAACAA 70546

864 TGTGTATGACACAGCTTAATATGCTCAATTTA 894

70547 TGAAGTTCTACATTAATACAGAGACAGATCA 70577

SOURCE

ORGANISM

Oryza sativa

Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

1 Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,

Katagiri, F., Quan, S., Tao, Y., Whitman, S., Xie, Z., Zhu, T. and Zou, G.

Plant genes involved in defense against pathogens

Parent: WO 0300898-A 5263 03-JAN-2003;

Syngenta Participations AG (CH)

FEATURES

source

1..2000

/organism="Oryza sativa"

/mol_type="unassigned DNA"

/db_xref="taxon:4530"

ORIGIN

Query Match 3.5%; Score 45; DB 6; Length 2000;

Best Local Similarity 8.4%; Pred. No. 14;

Matches 53; Conservative 290; Mismatches 285; Indels 1; Gaps 1;

189 GGAATATCATCATTCATCAATACAGAGAGCCAGAGAGTTCAGTACGATTAATCC 248

383 SRRKRKCMRKRKYKRMGYSRMSCKRARMKCRSRARWKGRCGTCRMSYGM 442

249 ATTTTGAAGGACATATTCATTCATGCTTTAGCTTATATTAACAGCAGGAGCATTTGA 308

443 RKSWMRMSKTKMMSRMRMRKCKSRITTMGKTRGCMGTGRCRYKRGSKMRKCR 502

309 TCTTGAATCATCATTCATTCATCAGACCTGATATCATTCACAAGAGATGAAACCA 368

503 RRMGRMYRMRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYR 562

369 GACATCCTCGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 428

563 YYYKMMYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRY 622

429 AAGAGCCGATTAAGTTCATCGGAGAGATTAAGTTCAGTTCAGTTCAGTTCAGTTC 488

623 CASKSKSAKRYAMMGATSGSRMSKSKSTCYWRKMSKSKSTCYWRKMSKSKSTCY 682

489 CCATTTTATCCCAATGTTATCAGATCAGATTAAGATATTTTGCACAGCTGCACAG 548

683 WRRYRAMCMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYR 742

549 TGCACATG-TTATGCCAAAAAGATRGATCTGGGTTGATGTTGCACTGCAATTT 607

743 YRSMMYYSMMWAKTKMRRRYATMMWYYSMMWYYSMMWYYSMMWYYSMMWYYS 802

608 ATGCTCATGTTATATTAAGATTTCAAGCCGCTTGAATAATGACGTTTCAGGTT 667

803 WTKCTTMYWATYWTGTWAAAMMAKTYMRGATGATRGARARAYWMAATWATAR 862

668 TGAAGAGTATCTGAGAGATCCCAACAGAGTGAAGAAATTTGAAATTAAGTGA 727

863 WMTKKGKAMATWMAKMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYR 922

728 AATTCAACATGAAGATGATACATTCATCAGATCAAGTATTAATGAGTACGTA 787

923 KGSAXMMWMMWKGKRGKWTYWTYCTTWMCAAGATATKWCAGAMWMSYSWTRYMRT 982

788 AGGGTGGCTCAGAAACACCAATTTGTA 816

983 WRMWMSRTAKRMARMMWKTAWMSKYA 1011

RESULT 39

LOCUS

DEFINITION

ACCESSION

VERSION

AR450463

Sequence 1127 from patent US 6673910.

AR450463

GI:42679520

2000 bp DNA linear PAT 22-MAR-2003

AX655393

AX655393

AX655393.1 GI:29158207

AX655393

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 26, 2005, 04:40:39 ; Search time 3059 Seconds
(without alignments)
15474.080 Million cell updates/sec

Title: US-10-069-062-6
Perfect score: 1299
Sequence: 1 atgcacaaagcatttgcgc.....aagactacatagcttataa 1299

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:
2: gb_est2:
3: gb_hic:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_est1:
9: gb_gsa2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	150.6	11.6	1059	CNS06MRJ	AL405845 T3 end of
C 2	131.8	10.1	920	CNS06MRJ	AL419103 T3 end of
C 3	71.8	5.5	973	CNS07ANQ	AL417100 T7 end of
C 4	65.6	5.1	917	CNS06MRJ	AL418769 T3 end of
C 5	63.8	4.9	920	CF824043	CF824043 EST701425
C 6	62.2	4.8	846	CO014330	CO014330 EST802665
C 7	62.2	4.8	922	CO008271	CO008271 EST796606
C 8	60.6	4.7	488	AZ931444	AZ931444 474_dh284
C 9	60	4.6	482	BZ304881	BZ304881 KD3133.D1
C 10	55	4.2	466	AU268449	AU268449 AU268449
C 11	55	4.2	479	AU268448	AU268448 AU268448
C 12	53.4	4.1	717	BJ343297	BJ343297 BJ343297
C 13	53.4	4.1	1101	CNS0039G	AL063921 Drosophila
C 14	51.6	4.0	322	AU052799	AU052799 AU052799
C 15	49.2	3.8	360	CO142585	CO142585 EST837226
C 16	49.2	3.8	412	CO144615	CO144615 EST829286
C 17	49.2	3.8	412	CO142023	CO142023 EST836694
C 18	49.2	3.8	461	CO148810	CO148810 EST83863
C 19	47.8	3.7	703	BO862174	BO862174 OGC20F15.
C 20	47.8	3.7	775	BU004632	BU004632 OGC5L17.Y
C 21	47.6	3.7	447	CO152583	CO152583 EST827647
C 22	47.6	3.7	475	CO137286	CO137286 EST831967
C 23	47.6	3.7	478	CO148500	CO148500 EST823553
C 24	47.6	3.7	561	CO136871	CO136871 EST831542

25	47.6	3.7	803	CO014331	CO014331 EST802666
26	47.6	3.7	1101	CNS0106X	AL098595 Drosophila
27	46.6	3.6	1101	CNS0182P	AL108811 Drosophila
28	46.2	3.6	716	BO866497	BO866497 OGC8C07.Y
29	46.2	3.6	745	BO870439	BO870439 OGC9A23.Y
C 30	46.2	3.6	1101	CNS017KX	AL108171 Drosophila
C 31	45.8	3.5	727	CG067021	CG067021 PUIDB42TD
C 32	45.4	3.5	499	CNS06D75	AL393447 T7 end of
C 33	45.2	3.5	715	BO860948	BO860948 OGC17C09.
C 34	45	3.5	697	BM165227	BM165227 EST567750
C 35	45	3.5	780	BM169836	BM169836 EST572359
C 36	45	3.5	805	BM161967	BM161967 EST564490
C 37	44.8	3.4	486	AQ400544	AQ400544 HS_5064.A
C 38	44.8	3.4	1101	CNS00DGI	AL069971 Drosophila
C 39	44.6	3.4	900	AZ549980	AZ549980 ENTDP94TF
C 40	44.6	3.4	1101	CNS00D07	AL069440 Drosophila
C 41	44.4	3.4	384	CO224433	CO224433 MS01021.B
C 42	44.4	3.4	705	BO869724	BO869724 QGD7A12.Y
C 43	44.4	3.4	802	CO122187	CO122187 GR_Eb03L
C 44	44.2	3.4	637	CR331656	CR331656 Medicago
C 45	44.2	3.4	752	BG585318	BG585318 EST487082

ALIGNMENTS

RESULT 1
CNS06MRJ/c 1059 bp DNA linear GSS 17-JUN-2001
LOCUS
DEFINITION T3 end of clone AU0AA006F09 of library AU0AA from strain CBS 3082
ACCESSION AL405845
VERSION AL405845
KEYWORDS
SOURCE
ORGANISM
GSS
Saccharomyces kluyveri
Saccharomyces kluyveri
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE
Autors
Soulier, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brotier, P., Casaregola, S., de-Montigny, V., Dujon, B., Durand, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nicolet, C., Wesolowski-Louvel, M., Wincker, P. and Weissenbach, J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
Autors
Neuvéglise, C., Bon, E., Lepingle, A., Wincker, P., Artiguenave, F., Galliard, C. and Casaregola, S.
Genomic exploration of the hemiascomycetous yeasts: 9. Saccharomyces kluyveri
FEBS Lett. 487 (1), 56-60 (2000)
MEDLINE
PUBMED
REFERENCE
Autors
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage, 2 rue Gaston Crémieux, CP 5706, 91057 Evry cedex, FRANCE. (E-mail: seqrefgenoscope.cns.fr - Web: www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactic var. lactic, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbicola, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of

COMMENT

the other extremity of this insert.

FEATURES
Location/Qualifiers

Source

1..1059

/organism="Saccharomyces kluyveri"

/mol_type="genomic DNA"

/strain="CBS 3082"

/cd_xref="taxon:4934"

/clone="AU0A006F09"

/note="end : T3"

/complement(<6..>824)

/note="similar to Saccharomyces cerevisiae ORF YMR220w [

ERG8 : phosphomethyltransferase kinase]"

/evidence=not_experimental

ORIGIN

Query Match 11.6%; Score 150.6; DB 9; Length 1059;
Best Local Similarity 51.9%; Pred. No. 5.6e-28;

Matches 426; Conservative 0; Mismatches 374; Indels 21; Gaps 3;

4 TCAGAAAGCATTTAGTGCACCTGGAAGACATTTCTGCTGGGATATTTGGTTCTGAG 63

815 TCAGAGCATTCAGGACACAGGTAAAGCCCTTAGCTGGTGAATCCTTGCTTAGAT 756

64 CCAATTATGATGCTTATGTCAGCATGTCATCAGATGATGATGATGATGATGAT 123

755 CCAATATATGATGCTTATGTCAGCATGTCATCAGATGATGATGATGATGATGAT 696

124 AAGAGAACCATTTGAAAGAAATCTAGATCAAAA-----TTCTTCAACCCCAATTTGA 177

695 AATGATTTGAAAGAAATCAAGATGATGATGATGATGATGATGATGATGATGAT 177

178 AAGGAGATGGAATATCATATCATATCATATCATATCATATCATATCATATCATAT 636

635 GATGCTGGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 237

238 CGCATATATCATTTTATGAGGACATATATATATATATATATATATATATATATAT 576

575 CTGGAATATCATTTTATGAGGACATATATATATATATATATATATATATATATAT 297

298 GAAGCATTTGATCTGGAATCATATATATATATATATATATATATATATATATATAT 516

515 TCGACATGATCTCAAT 357

358 ACTGAAACCAAGATCTTGAATGAGAAATATATATATATATATATATATATATATAT 456

455 GAGGGAAGTCTCCGGAAGAAATGTTCAAAATTTAATTTCAATTAAGAACCAATC 417

418 ACCGAAGTGAAGAAAGCCGATTTAGTTCATGCGAGATTAAGTGTCAAGTTGCCACA 396

395 ACGAAGTCCCAAGCTGGGCTGGGCTCTTCTGCGGATTTGGTAACAGTTCTAAGTGT 477

478 AGTTTATATCCATTTATCCCATATGTTATCATGATGATGATGATGATGATGATGAT 336

335 GGTTTATATCCATTTATCCCATATGTTATCATGATGATGATGATGATGATGATGAT 528

529 TTGCAACAGTTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAG 276

275 ATTCAATATATGCAAGATTTCTCATTTGATGATGATGATGATGATGATGATGATGAT 588

588 GATTTGCAATGCAATTTATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 216

215 GATTTGCGGAGCCACATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 648

649 AATGACGTTTACGTTCTAGAAAGTATCTGGAAGTTCCTCCACAGAGTTGAAAGAAA 156

155 AATATTTTACACCTGCTGATTTCTGTTGAATACCAAGATTCCTCGTGAAGTTG----- 708

709 TTGATTTGAAGTAACTGGGAATTCAGAAAGATTAAGATTAAGATTAAGATTAAGATTA 102

101 GTTATGAAATATGATTTGAAATATCAAAATGAAGATTAAGATTAAGATTAAGATTAAG 768

769 TTATTTATGGGTGACGTTCAAGAGGTGCTCAGAAACACCA 809

Db 41 TTATGTCATGGAGATGTTAAACACCGCTTCTGAAACTACCA 1

RESULT 2 CDS06WZT 920 bp DNA linear GSS 06-JUL-2001

LOCUS T3 end of clone AX0A036H05 of library AX0A from strain CBS 7064

DEFINITION of Pichia farinosa, genomic survey sequence.

ACCESSION AL419103.1 GI:12202281

VERSION GSS.

KEYWORDS Pichia farinosa

SOURCE Pichia farinosa

ORGANISM Pichia farinosa

Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetaceae; Saccharomycetaceae; Pichia.

1 (bases 1 to 920)

Soulier, J. L., Aigle, M., Artiguenave, F., Blandin, G.,

Bolotin-Fukuhara, M., Bon, E., Brotier, P., Casaregola, S.,

de Montigny, J., Dujon, B., Durand, P., Lepingle, A., Llorente, B.,

Malpertuy, A., Neuvéglise, C., Ozier-Kalogiropoulos, O., Potier, S.,

Saurin, M., Tekala, F., Toffano-Nioche, C., Weolowski-Louvel, M.,

Wincker, P., and Weissenbach, J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of

yeast species for molecular evolution studies

FEBS Lett. 487 (1), 3-12 (2000)

REFERENCE

2 (bases 1 to 920)

de Montigny, J., Spehner, C., Soulier, J., Tekala, F., Dujon, B.,

Wincker, P., Artiguenave, F., and Potier, S.

Genomic exploration of the hemiascomycetous yeasts: 15. Pichia

sorbitophila

FEBS Lett. 487 (1), 87-90 (2000)

JOURNAL

20584725

PUBMED

11152890

3 (bases 1 to 920)

Genoscope.

Direct Submission

Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage,

2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :

seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)

This GSS is part of a random genomic sequencing program of thirteen

yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces

Saccharomyces kluyveri, Zygosaccharomyces rouxii,

lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia

angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,

5 kb were prepared and both extremities were sequenced. See

keywords for description of this sequence and for the sequence of

the other extremity of this insert.

Location/Qualifiers

1..920

/organism="Pichia farinosa"

/mol_type="genomic DNA"

/strain="CBS 7064"

/db_xref="taxon:4920"

/clone="AX0A036H05"

/clone_1kb="AX0A"

/note="end : T3"

<3..>893

/note="similar to Saccharomyces cerevisiae ORF YMR220w [

ERG8 : phosphomethyltransferase kinase]"

/evidence=not_experimental

ORIGIN

Query Match 10.1%; Score 131.8; DB 9; Length 920;

Best Local Similarity 54.1%; Pred. No. 4.3e-23;

Matches 349; Conservative 0; Mismatches 272; Indels 24; Gaps 3;

335 CTGATATATATCAAGAAAGATTAAGAACCAACATCTCTCAATGAGAGAAAAACAT 394

Db 1 CTGGGTACCATACTAGATGACACGACCGGTGTCTCAAGCAACGGCAACGGCAAT 60
 QY 395 TTCTTACCATTCCTGTCGTCATTACCGAAGTGGAAAAGACCGGATTAGTTCACTGGCAG 454
 Db 61 TCTTATACCATTAAGAAAGAAATCAGACGAGTTGCAAAAACAGACTTGGGTGCTGGCCG 120
 QY 455 GATTAGTGTCACTGTTGGCACAAGTTTATATCCCAATTTATCCCA---TGTATCA 511
 Db 121 GATTGTGTGAGTAAGACGGCGGCACTGCTCTTACTTCACTGCAACCGGAGCTTG 180
 QY 512 GTACGAATAAATATTTTGGACACAGTTGACAGATTTGACATTTATATGCCCAAAA 571
 Db 181 ATTCTACCTCGCAACCGGTCCATTAATGACGAGATTAATCACTGTACCGCCCAAAA 240
 QY 572 ACATAGGATCTGGGTTGATGTTGCACTGCAATTTATGATGTTATATAGAAAT 631
 Db 241 AATCGGCTCTGTTTCACTGTTGACAGCTGTGTACGCTCTATTTGTTACCGCCGTT 300
 QY 632 TTTCAGCCAGCTTGA-----TAAATGACGCTGTTGACGTTCTAGAAAGTATC 679
 Db 301 TCCAGCTAGCTTGAATCGACGAATATCTAAGACAGAGTTTCAAGTATGCAAGTGA 360
 QY 680 CTGAG-----AAGTCCCAACAGAGTTGAAAATTTGATGAAAGTAACTGGGAAT 730
 Db 361 CTACGCTAGCGCTGTGACTCAAGAACTCTCTACAGACATAGTGGGTTCTGACTGG 420
 QY 731 TCAACATGAAGATGTATACATTAAGCAATAGGAATCAAGTTAATAGGTTGACGTTCA 790
 Db 421 TCAAAATGAATAATGTGCTTGCCTGCGCGCGGCTGACACTCGTAATGGAGACGTTCT 480
 QY 791 GTGGCTCAGAAAACCCCAATTTGTATACAGAGTACTCCATGCAAAAAGAAAAGCCAG 850
 Db 481 TCGGTAGGAGACGCGCAATTTGTATGCTATGCTCAAGTGAAGCAAGAAACCTGG 540
 QY 851 AAGAAAGCTCTGTTGTATGACACAGCTTAATATGTCCTAATTTACGTTTATGAAG 910
 Db 541 TCGAGGGCAGAAATCAATATTGAGAGCTTAATAAGCCCAATGGTTCATCGCGCAT 600
 QY 911 TGAGGGAATGCGTGAATAATCACTGACGCCAGACTTATA 955
 Db 601 TATCTACTGCTTACGAACTCCAGCGCGCAACGAGCGGCTTACA 645

RESULT 3
 Locus: CNS07AVQ/c 973 bp DNA linear GSS 08-JUL-2001
 Definition: T7 end of clone BC0A007C01 of library BC0A from strain CBS 767 of Debaryomyces hansenii, genomic survey sequence.
 Accession: AL437100
 Version: AL437100.1 GI:12220513
 Keywords: GSS.
 Source: Debaryomyces hansenii (anamorph: Candida famata)
 Organism: Debaryomyces hansenii
 Reference: Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Debaryomyces.
 Authors: Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saulin,W., Tekala,F., Toffano-Nicolas,C., Wesolowski-Louvel,M., Winkler,P. and Weissbach,J.
 Title: Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
 Journal: FEBS Lett. 487 (1), 3-12 (2000)
 Medline: 11152876
 Pubmed: 20584724
 Reference: 2 (bases 1 to 973)
 Authors: Lepingle,A., Casaregola,S., Neuvéglise,C., Bon,E., Nguyen,H., Artiguenave,F., Winkler,P. and Gallardin,C.
 Title: Genomic exploration of the hemiascomycetous yeasts: 14. Debaryomyces hansenii var. hansenii
 Journal: FEBS Lett. 487 (1), 82-86 (2000)

MEDLINE 20584724
 PUBMED 11152876
 REFERENCE 3. (bases 1 to 973)
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxi, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
 Location/Qualifiers
 1..973
 /organism="Debaryomyces hansenii"
 /mol_type="genomic DNA"
 /strain="CBS 767"
 /variety="hansenii"
 /db_xref="taxon:4959"
 /clone="BC0A007C01"
 /clone_1b="BC0A"
 /note="end : 77"
 complement(432..>139)
 /note="similar to Saccharomyces cerevisiae ORF YMR220w [ERG8 ; phosphomevalonate kinase]"
 /evidence=not_experimental
 <278..>972
 /note="similar to Saccharomyces cerevisiae ORF YBR166c [TYR1 ; prephenate dehydrogenase (NADP+)]
 1 putative frameshift(s)"
 /evidence=not_experimental

ORIGIN
 Query Match 5.5%; Score 71.8; DB 9; Length 973;
 Best Local Similarity 74.0%; Pred. No. 1.9e-07;
 Matches 91; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 5 CAAAGCATTTAGTACCTGGAAGAAAGATTTCTGCTGATGATTTGTTCTTGAGC 64
 Db 141 CTAGAGCATTCCTGCTCTGGAAGAAAGATTTTGGCGGTGATACCTCGTATAGAC 82
 QY 65 CAATTATGATGCTTATGTGACAGCATTTGTCATCAGAAATGATGCAATTATACACCA 124
 Db 81 CAATATACATCTTATGTGACAGCGCTTTCTTAAGAAATGATGCAATTGTTAAAAATA 22
 QY 125 AAG 127
 Db 21 GAG 19

RESULT 4
 Locus: CNS06WQJ 917 bp DNA linear GSS 06-JUL-2001
 Definition: T3 end of clone AX0A034D03 of library AX0A from strain CBS 7064 of Pichia farinosa, genomic survey sequence.
 Accession: AL418769
 Version: AL418769.1 GI:12201829
 Keywords: GSS.
 Source: Pichia farinosa
 Organism: Pichia farinosa
 Reference: Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.
 Authors: Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,


```

Db      384 ATTGAACCTGAGTGCAGACAGAAATTATTGAATCCCTTTTGAAAATTGACGGGGTCATT 325
Qy      1111 GGTGCTGTGCTTCCAGAGTCTGTGATACATGACATAGCTGTATTAGTTGGAAT 1170
Db      324 GGTGCGCTTTTCCCGCGCGGCGGATATGATGACCTGTGCTTGTGTAAGATGAT 265
Qy      1171 C 1171
Db      264 C 264

RESULT 6
C0014330/c      846 bp      mRNA      linear      EST 09-JUN-2004
DEFINITION      EST02655 Coccidioides posadasii spherule cDNA library, 0.4 to 2.3
                  kb Coccidioides posadasii cDNA clone CIEC383 3' end, mRNA sequence.
ACCESSION      C0014330.1 GI:48521219
VERSION      C0014330.1 GI:48521219
KEYWORDS
SOURCE      EST.
ORGANISM      Coccidioides posadasii
                  Coccidioides posadasii
                  Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
                  Onygenales; Mitosporic Onygenales; Coccidioides.
REFERENCE      1 (bases 1 to 846)
                  Gardner, M.J. and Cole, G.T.
                  Analysis of gene expression in Coccidioides posadasii mycelia and
                  spherules via expressed sequence tags
                  Unpublished (2003)
JOURNAL
COMMENT      Other_ESTs: EST802666
                  Contact: Gardner MJ
                  The Institute for Genomic Research
                  9712 Medical Center Drive, Rockville, MD 20850, USA
                  Tel: 301 838 3519
                  Fax: 301 838 0208
                  Email: gardner@tigr.org.

FEATURES
source      1..846
                  /organism="Coccidioides posadasii"
                  /mol_type="mRNA"
                  /strain="C735"
                  /db_xref="taxon:199306"
                  /clone="CIEC383"
                  /dev_stage="spherules"
                  /lab_host="E. coli DH10B, T1 phage resistant"
                  /note="Vector: pExpress 1; Site_1: Not I; Site_2: Eco RV;
                  Coccidioides posadasii spherule cDNA library, 0.4 to 2.3
                  kb"

ORIGIN
Query Match      4.8%; Score 62.2; DB 7; Length 846;
Best Local Similarity 46.7%; Pred. No. 6e-05;
Matches 309; Conservative 0; Mismatches 313; Indels 39; Gaps 2;

Qy      523 GATATTTTGCACACGTTGCACAGATTGACATTGTTATGCCCAAAAAAGATGAGATCT 582
Db      821 GATTAACATACATACCTGCTCAAGTCTCACTGTGCGGCCCAAGTAAAGTGGCTTT 762
Qy      583 GGGTTTGATGTTGCAACCTGCAATTATATGCTCTGATTTATATAGAAATTTCAAGCAAGCT 642
Db      761 GGAATTTGACATTGCGGACGCGTACGATCTTGTCTTTACCGTGTTCACACATCT 702
Qy      643 TTGATTAATGACGTGTTTCAAGTCTTCAAGAAAGTACCTCGAAGATTCCCAAC----- 695
Db      701 GTTCTAGGACCTGGCGGAGGTGGCTCCCGCACTTTTGAACAGCGCTATTATACAGTT 642
Qy      696 -----AGAGTTGAAAAAATTGATTGAAAGTAAGTGGGAATTCAAACATGAAGATGTACA 750
Db      641 GTTGAAGACCTCAATACCAAAAACTTTGGACACAGAGTGTGTAGACTTTGATTTAA 582
Qy      751 TTACCAATACGAATCAAGTTATTAATGATGACGTCAAGGGTGGCTCAGAAAACCCCAA 810

```

```

Db      581 CTCACAGGGGGCTCAAAATGCTCTTTGCCAGCTGAGTGGCTTCCAGACCCCCGGC 522
Qy      811 TTGATATCACAGATCACTCCATGATGAAGAAAAAGCCAGAAAGATCTGTGTGTAT 870
Db      521 ATGTTTAAAGATCTCTGAGTGTGGGGAACGAACCCGGGAGCGCAGAGTGTCTGG 462
Qy      871 GACGAGCTTAATATGTCGAATTTACGTTTATGAAAGAAATGAGGAAATGCGTAAGAA 930
Db      461 ACGGGGGTACAGCGGAACATGAAGAAAGCTTGCCTCAGCTCAACGATTTGACACAAAC 402
Qy      931 TACGACTACAGACCCAGAACTTATTTAAAGATTAGATCACTCTGTGAGCTTTGACT 990
Db      401 CGCAAC-----GTATCCAGAACTATGATGACTCAAC 369
Qy      991 GTTGGATTAAGACATCAGAAAGGTTATACAGATTAAACAAAAATCAGAGTTCCA 1050
Db      368 AACTGATACATCCGACATGAGATGTGGAATTAGTCAATGACAGAGAAATGCGGGTTCCC 309
Qy      1051 ATTGAACCTGATGTCGAACCCAGTTGTTGACCGCTTCAAGAGATTCTGTGTGTT 1110
Db      308 ATTGAACCTGAGTGCAGACAGAAATTATGAATGCTTTTGGAAATTGACGGGGTCATT 249
Qy      1111 GGTGCTGTGCTTCCAGAGTCTGTGATACATGACATAGCTGTATTAGTTGGAAT 1170
Db      248 GGTGCGCTTTTCCCGCGCGGCGGATATGATGACCTGTGCTTGTGTAAGATGAT 189
Qy      1171 C 1171
Db      188 C 188

RESULT 7
C0008271/c      922 bp      mRNA      linear      EST 09-JUN-2004
DEFINITION      EST976606 Coccidioides posadasii spherule cDNA library, 0.4 to 2.3
                  kb Coccidioides posadasii cDNA clone CIEB521 3' end, mRNA sequence.
ACCESSION      C0008271
VERSION      C0008271.1 GI:48515160
KEYWORDS
SOURCE      EST.
ORGANISM      Coccidioides posadasii
                  Coccidioides posadasii
                  Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
                  Onygenales; Mitosporic Onygenales; Coccidioides.
REFERENCE      1 (bases 1 to 922)
                  Gardner, M.J. and Cole, G.T.
                  Analysis of gene expression in Coccidioides posadasii mycelia and
                  spherules via expressed sequence tags
                  Unpublished (2003)
JOURNAL
COMMENT      Other_ESTs: EST96607
                  Contact: Gardner MJ
                  The Institute for Genomic Research
                  9712 Medical Center Drive, Rockville, MD 20850, USA
                  Tel: 301 838 3519
                  Fax: 301 838 0208
                  Email: gardner@tigr.org.

FEATURES
source      1..922
                  /organism="Coccidioides posadasii"
                  /mol_type="mRNA"
                  /strain="C735"
                  /db_xref="taxon:199306"
                  /clone="CIEB521"
                  /dev_stage="spherules"
                  /lab_host="E. coli DH10B, T1 phage resistant"
                  /note="Vector: pExpress 1; Site_1: Not I; Site_2: Eco RV;
                  Coccidioides posadasii spherule cDNA library, 0.4 to 2.3
                  kb"

ORIGIN
Query Match      4.8%; Score 62.2; DB 7; Length 922;
Best Local Similarity 46.7%; Pred. No. 6.1e-05;

```

Matches 309; Conservative 0; Mismatches 313; Indels 39; Gaps 2;

```

QY 523 GATATTTTGCACAACTTGCACAGATTCGATTTTATGCCCCAAAAAGATGATCT 582
DB |||||
DB 836 GATTAACATACATACCTGCTTACCTCTCAGCTGCGCGCAAGTAAGAGCTCT 777
QY 583 GGGTTGATGCTGCACTGCAATTTATGCTGATTTGATTAAGAAATTTCCAGCT 642
DB |||||
DB 776 GATTTGACATTTGCGCGCAGCGTGAACGATCTTTCATCCGTTTCACCATCT 717
QY 643 TTGATTAATGACGTTTTCAGTTCTGAAAGATATCTTGAAGTTCCAC----- 695
DB |||||
DB 716 GTTCTCAGCGACCTGGGAGGTTGCTCCCCGAGTTTGAACGCGCTTATTAAGTT 657
QY 696 -----AGAGTTGAAAAATTTGATTGAAGTAACTGAGAAATTCAAACATGAAGATGACA 750
DB |||||
DB 656 GTTGAAGCTCTCAATTAACGAAAACTTTGGGACAGAGATGTGTACCTTGGATTTAA 597
QY 751 TTACCATACGGAATCAATTAATTAATGGTGAAGCTCAAGGTTGCTCAAGAAACCCCAA 810
DB |||||
DB 596 CTCCCAAGGGGCTCCAAATGCTCTTTGCGAGCTGACGCGGTTCCAGACCCCGGC 537
QY 811 TTGGTATCAGATTACTCCATGAAAAAGAAAAAGCAAGAAAGCTCTGTTGTAT 870
DB |||||
DB 536 ATGTTTAAAAAGGTGTGAGTGCGGGAACAGAACCCGGGACCGACGCTCTGTGG 477
QY 871 GACCACTTAATAGTCCCAATTTACGTTTATGAAGAAATTAAGGAAATGCGTGAATA 930
DB |||||
DB 476 ACCGGGCTACAGGGAACATGAATAAGCTTGCGCTCGACATTCAGCAATTTGCAAAAC 417
QY 931 TACGACTCAGACCCAGACGCTTAATTAAGAGTTAGTCACTGCTGTTGAGCCTTGACT 990
DB |||||
DB 416 CGC-----AATGATGCCAGATTTATGAATCTCAC 384
QY 991 GTTGGATTAAGACATCAGAAAGGTTTACACATTAACAAATTAAGAGTTTCA 1050
DB |||||
DB 383 AACTTGATCACTCGCAGTGAATGATTTAGTCAATGAAGAAATGCGGGGTTCC 324
QY 1051 ATTGAACCTGATTCCTCAAACTGTTTGAACGTTGTCAGAGATTCCTGTTGTGT 1110
DB |||||
DB 323 ATTGAACCTGAGTGAGCAAGATTAATTAATGATTCCTTTCCAAATTAAGCGGCTCAT 264
QY 1111 GTTGTGTGTTTCCAGTGTGCTGATGATGATGATGATGATGATGATGATGATGAT 1170
DB |||||
DB 263 GGTGGCGTTGTTCCGCGCGCGCGGCGGATGATGATGATGATGATGATGATGAT 204
QY 1171 C 1171
DB 203 C 203

```

RESULT 8
LOCUS A2931444 488 bp DNA linear GSS 01-APR-2001
DEFINITION 474.dh284b12.s1 Saccharomyces unisporus NRRL Y-1556 Saccharomyces unisporus genomic clone 474.dh284b12.s1, genomic survey sequence.
ACCESSION A2931444
VERSION A2931444.1 GI:13502355
KEYWORDS GSS.
SOURCE Saccharomyces unisporus
ORGANISM Saccharomyces unisporus
REFERENCE 1 (bases 1 to 488)
AUTHORS Gish,W.R., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis
JOURNAL Unpublished (2001)
COMMENT Department of Genetics
Washington University Medical School
Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA

FEATURES
source
1..488
/organism="Saccharomyces unisporus"
/mol_type="genomic DNA"
/strain="NRRL Y-1556 (CBS 398)"
/db_xref="taxon:27294"
/clone="474.dh284b12.s1"
/clone_lib="Saccharomyces unisporus NRRL Y-1556"
/note="Random genomic sequence"

ORIGIN

Query Match 4.7%; Score 60.6; DB 8; Length 488;
Best Local Similarity 54.7%; Pred. No. 0.00033;
Matches 168; Conservative 0; Mismatches 129; Indels 10; Gaps 2;

```

QY 991 GTTGGATTAAGACATCAGAAAGGTTTACAAACATTAACAAATTAAGAGTTTCA 1050
DB |||||
DB 5 GCTCAATTAACAAATTAAGAGATTAATTAATTAATTAATTAATTAATTAATTAAT 64
QY 1051 AT-TGAACCTGATGCCAAACCCAGTTGTTGACCGTGTCAAGATTTCTGTTGT 1109
DB |||||
DB 65 ATGTGAACCAATGTGTCAACCAATTTGCTTAACATTTGTTGACTTAATGTTGAT 124
QY 1110 TGTGTGTGTGTTCCAGGTGCTGTGTGATGATGATGATGATGATGATGATGATGAT 1169
DB |||||
DB 125 AGGTGTGTGTCTCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 184
QY 1170 TCAAGTGGAAATTTTAAACAGAAACTCTTGAATAATCAATTAATTTTCTAATGTT 1229
DB |||||
DB 185 -----AATGATTAATTAAGATTAATTAAGATTAATTAAGATTAATTTTCAAAATGTCAC 235
QY 1230 CTGGGTTGATTTTGAAGGCAACGAAAGGTGTCTTGAAGAAAAACAGAGCTATAT 1289
DB |||||
DB 236 TTGATGATGATGATTAACAAACAAATTTGATTAAGAAAGAAATTCAAACATTTATGA 295
QY 1290 AGGTTTA 1296
DB |||||
DB 296 AGATCTTA 302

```

RESULT 9
LOCUS B2304881/c 482 bp DNA linear GSS 31-OCT-2002
DEFINITION KD3193.pl Kluyveromyces delphensis Random Genomic Library
KEYWORDS Kluyveromyces delphensis genomic clone KD3193, genomic survey
SEQUENCE.
ACCESSION B2304881
VERSION B2304881
KEYWORDS GSS.
SOURCE Kluyveromyces delphensis
ORGANISM Kluyveromyces delphensis
REFERENCE 1 (bases 1 to 482)
AUTHORS Wong,S., Fares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.
TITLE the 'asexual' pathogenic yeast Candida glabrata
JOURNAL Genome Biol. 4 (2), K10 (2003)
MEDLINE 22508158
PUBMED 12620120

COMMENT Contact: Wong S
Department of Genetics, Smurfit Institute
Trinity College Dublin
Dublin 2, Ireland
Tel: 353 1 6082319
Fax: 353 1 6798558
Email: swong@tcd.ie
Class: plasmid ends.
Location/Qualifiers

FEATURES

```

source
1..482
/organism="Kluyveromyces delphensis"
/mol_type="genomic DNA"
/strain="CBS 2170"
/db_xref="taxon:51657"
/clone="KD3193"
/clone_1ib="Kluyveromyces delphensis Random Genomic
Library"
ORIGIN

Query Match 4.6%; Score 60; DB 8; Length 482;
Best Local Similarity 61.5%; Pred. No. 0.0019;
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 993 TGGCATTAAGAACACAGAAAGGTTACACATTAAACAAAATAGAGGTTCCAT 1052
Db 272 TGCATTTAAACAGATTAGAAAACCTTCAGATTAAACAGAAATCAGGTCGAAT 213
Qy 1053 TGAACCTGATGCCAACCCAGTTGTGACCGTTGCAGAGATTCTCTGTTGTTGG 1112
Db 212 CGAACCAAGAACACAACTAGGTTACTGATGCATGTTCCAAATTGAATGGGTATGG 153
Qy 1113 TGGTGTGTTCCAGTCTGCTGTGATACGATCAAT 1148
Db 152 AGCGCTCGTTCAGAGTGGTGGTTATGACGCTAT 117

RESULT 10
LOCUS AU268449 466 bp mRNA linear EST 26-APR-2004
DEFINITION AU268449 VS Dictyostelium discoideum cDNA clone VS1284 3', mRNA
sequence.
ACCESSION AU268449
VERSION AU268449.1 GI:20527247
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
1 (bases 1 to 466)
Urushihara, H., Morio, T., Saito, T., Kohara, Y., Koriki, E., Ochiai, H.,
Maeda, M., Williams, J.G., Takeuchi, I. and Tanaka, Y.
Analyses of cDNAs from growth and slug stages of Dictyostelium
discoideum
Nucleic Acids Res. 32 (5), 1647-1653 (2004)
JOURNAL Contact: Hideko Urushihara
COMMENT Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp.

FEATURES
source
1..466
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="VS1284"
/sex="mat A"
/dev_stage="vegetative"
/clone_1ib="VS"

ORIGIN

Query Match 4.2%; Score 55; DB 1; Length 466;
Best Local Similarity 52.4%; Pred. No. 0.0038;
Matches 121; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

Qy 957 TAAAGAGTTAGATCTTCTGTGAGCCTTGTGACTGTGGATTAGAACATCAGAAAG 1016
Db 98 TAAATATTTGGTAAACCTTTAATTAATTAAGAAATCATTTTAAAGATTAAGATCAT 157
Qy 1017 GTTACAGCATTTACACAAATTCAGAGGTTTCAATTGAACCTGATGCCAACCGATT 1076
Db 1017 GTTACAGCATTTACACAAATTCAGAGGTTTCAATTGAACCTGATGCCAACCGATT 1076

```

```

Db 158 AATGAGAGAAATGGGTAAATTTCTGATGATCAATTCGAAACCAATTGAACCAATGAAAT 217
Qy 1077 GTTGACCGTTGTCAAGAGATTTCTGTTGTGGTGGTGGTCCAGGTCGTGG 1136
Db 218 GCGTATATATACATGATCATATCATGATGATTTATTTCTGCTGATGATGATGATGATG 277
Qy 1137 ATACGATCAATAGCTGTTATTTAGTTGGAAATCAATGAGAAATTTTAA 1187
Db 278 TTTGATGCTTTATTCGCTATGCTTATCATCATCAATCAAGATGAATTAATAA 328

RESULT 11
LOCUS AU268448 479 bp mRNA linear EST 26-APR-2004
DEFINITION AU268448 VS Dictyostelium discoideum cDNA clone VS1284 5', mRNA
sequence.
ACCESSION AU268448
VERSION AU268448.1 GI:20527246
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
1 (bases 1 to 479)
Urushihara, H., Morio, T., Saito, T., Kohara, Y., Koriki, E., Ochiai, H.,
Maeda, M., Williams, J.G., Takeuchi, I. and Tanaka, Y.
Analyses of cDNAs from growth and slug stages of Dictyostelium
discoideum
Nucleic Acids Res. 32 (5), 1647-1653 (2004)
JOURNAL Contact: Hideko Urushihara
COMMENT Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp.

FEATURES
source
1..479
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="VS1284"
/sex="mat A"
/dev_stage="vegetative"
/clone_1ib="VS"

ORIGIN

Query Match 4.2%; Score 55; DB 1; Length 479;
Best Local Similarity 52.4%; Pred. No. 0.0038;
Matches 121; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

Qy 957 TAAAGAGTTAGATCTTCTGTGAGCCTTGTGACTGTTCGATTAAGAACATCAGAAAG 1016
Db 86 TAAATATTTGGTAAACCTTTAATTAATTAAGAAATCATTTTAAAGATTAAGATCAT 145
Qy 1017 GTTACAGCATTTACACAAATTCAGAGGTTTCAATTGAACCTGATGCCAACCGAGTT 1076
Db 146 AATGAGAGAAATGGGTAAATTTCTGATGATGATGATGATGATGATGATGATGAT 205
Qy 1077 GTTGACCGTTGTCAAGAGATTTCTGTTGTGGTGGTGGTCCAGGTCGTGG 1136
Db 206 GCGTATATATACATGATCATATCATGATGATGATGATGATGATGATGATGATGAT 265
Qy 1137 ATACGATCAATAGCTGTTATTTAGTTGGAAATCAATGAGAAATTTTAA 1187
Db 266 TTTGATGCTTTATTCGCTATGCTTATCATCATCAATCAAGATGAATTAATAA 316

RESULT 12
LOCUS BU343297 717 bp mRNA linear EST 07-MAR-2002
DEFINITION BU343297 Dictyostelium discoideum cDNA library, AF Dictyostelium
discoideum cDNA clone dd22j18 3', mRNA sequence.

```

```

ACCESSION   BJ343297
VERSION     BJ343297.1  GI:19251660
KEYWORDS
SOURCE      Dictyostelium discoideum
ORGANISM    Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
REFERENCE   1 (bases 1 to 717)
AUTHORS     Ureshihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
TITLE       Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
JOURNAL     Unpublished (2002)
COMMENT     Contact: Tadao Shin-I
            Center For Genetic Resource Information
            National Institute of Genetics
            111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES
    source
        location/Qualifiers
            1..717
            /organism="Dictyostelium discoideum"
            /mol_type="mRNA"
            /strain="AX4"
            /db_xref="taxon:44689"
            /clone="dda22j18"
            /sex="mat A"
            /dev_stage="aggregation stage"
            /clone_idb="Dictyostelium discoideum cDNA library, AP"
ORIGIN
Query Match
Best Local Similarity 4.1%; Score 53.4; DB 4; Length 717;
Matches 120; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
QY 957 TAAAGCTTGATCATCTGCTGAGCCCTTGACCTGCTGCAATTAAGACATCAGAAAG 1016
DB 377 TAATTAATATGGTAAACCTTAATATCAATTAAGATCAATTTTAAAGTAAATGATCAT 318
QY 1017 GTTCAAGCATTACACAAATATCAGAGTTCCATTTGAATGATGTCACAAACCAAGT 1076
DB 317 AATGAGAGAAATGGTAAATTTGCTGATGATGATCAATGGAACCAATTGAACCAATGAAAT 258
QY 1077 GTTGAACGGTGTCAAGAGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1136
DB 257 GCGTATCATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 198
QY 1137 ATAGATGCAATAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1187
DB 197 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 147
RESULT 13
CNS00396/c 1101 bp DNA linear GSS 03-JUN-1999
LOCUS
DEFINITION BACRO8K10 of RBC1-98 library from Drosophila melanogaster (fruit
VERSION AL063921
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).

```

The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using the BACs. For further information please see <http://www.fruitfly.org>. The BDGP *Drosophila melanogaster* BAC library was prepared by Kazuo Oosawa and Aaron Mammiger in Peter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RBC1-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the pi and Bsr I libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```

FEATURES
    source
        location/Qualifiers
            1..1101
            /organism="Drosophila melanogaster"
            /mol_type="genomic DNA"
            /db_xref="taxon:7227"
            /clone="BACRO8K10"
            /clone_idb="RBC1-98"
            /note="Tent : TET3"
ORIGIN
Query Match
Best Local Similarity 4.1%; Score 53.4; DB 9; Length 1101;
Matches 85; Conservative 258; Mismatches 219; Indels 3; Gaps 1;
QY 737 ATGAAAGATGACATTCATCGAATCAAGATTAATTAAGCTGACGTCAGGTCGCT 796
DB 1093 DTDMDTRDDDDMDWTKMTWMDRMDRMDRMDRMDRMDRMDRMDRMDRMDRMDRMD 1034
QY 797 CAGAAACACCAATTTGATCAGAGTCTCCATGGAAGAAAGAAAGCCAGAAAGAA 856
DB 1033 WDTWMDKMMWATAKDTTATMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMT 974
QY 857 GCTGCTGCTGATGACCAAGCTTAATGATGATGATGATGATGATGATGATGATGATG 916
DB 973 KKKDKKKDDDDDKKKKKKAAKAAKATMTMTMTMTMTMTMTMTMTMTMTMTMTMT 914
QY 917 AAATGCGTGAATAATGACCTCAGACCCAGAGCTTAATTAAGATTAATGATCATCTG 976
DB 913 DKDDGKKADDDDDDDDKKDDDKKDDDKKDDDKKDDDKKDDDKKDDDKKDDDKK 854
QY 977 TTGAGCCCTT--GACGTTGCGATTAAGACATCAGAAAGGTTTACAGCATTAAC 1033
DB 853 TWDAADDDWADRWAMAMKWDAMWAGARTADRDMDGAGRGARRRRRRKRRADDK 794
QY 1034 AAAAATCAGAGGTTCCATTTGAACCTGATGTCAAACCCAGTTGTCGACGTCGACG 1093
DB 793 RDAADRDADAATWTTTTRDDDDDKKTDVTWMAADRTWDRDDDDDRDRAATGRK 734
QY 1094 AGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 734
DB 733 WRTWRRKRRKRDITWDDADDDTDARDRRRRGGDGDAGKGTGKRRRRRRRTWDR 674
QY 1154 TATTAAGTGTGAAATCAAGTGAATTTTAAGCAAGAAACTCTGAAATTCAGATT 1213
DB 673 DAWMDAAWMTTDTTDDDDDKRRKRRRRRTTAADAAWMTWAMWAKWMDKTR 614
QY 1214 ATTTTCATTAATGTTTACCTGGGTTGATTAAGAGCAACAGAGTGTACTTAAAGAA 1273
DB 613 ADRDWRWADITWIDARADWDWAKAWARARARARARARARARARARARARARAR 554
QY 1274 AACCAAGACTATTAAGGTTTATA 1298
DB 553 RAAMWAMWAMWATTTATWTTTWTW 529
RESULT 14
LOCUS
DEFINITION A052799 322 bp mRNA linear EST 28-APR-1999
A052799 Dictyostelium discoideum SL (H.Ureshihara) Dictyostelium
discoideum cDNA clone SLF121, mRNA sequence.

```

ACCESSION	AU052799
VERSION	AU052799.1
KEYWORDS	GI:4701282
SOURCE	EST.
ORGANISM	Dicystostelium discoideum Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
REFERENCE	1 (bases 1 to 322)
AUTHORS	Morita,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,K., Mitra,B.N., Pl.M., Sato,T., Takemoto,K., Yasukawa,H., Williams,J., Mada,W., Takeuchi,T., Ochiai,I. and Tanaka,Y. Developmental cDNA in Dictyostelium discoideum Unpublished (1998) Contact: Hideko Urushihara Institute of Biological Sciences University of Tsukuba 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan Tel.: 81-298-53-4664 Fax: 81-298-53-6614 Email: hideko@biol.tsukuba.ac.jp PROJECT = Dicystostelium discoideum cDNA project in Japan.
TITLE	
JOURNAL	
COMMENT	Location/Qualifiers 1..322 /organism="Dicystostelium discoideum" /mol_type="mRNA" /strain="AX4" /db_xref="taxon:44689" /clone="SLF121" /dev_stage="slug" /clone_lib="Dictyostelium discoideum SL ('H.Urushihara)"
FEATURES	
source	
Query Match	4.0%; Score 51.6; DB 1; Length 322;
Best Local Similarity	58.4%; Pred.No.0.056;
Matches	90; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
Qy	1034 AAAAATCAGAGGTTCCATTGAACCTGATGCACAACCAGGTGGACCGTGTCAAG 10933
Db	4 AAATTCGATGATCACCATGGAAACCAATTGAACAAATGCAATGGCTATCATCAATGT 63
Qy	1094 AGATTCCGTGTTGTTGGTGATGCTGCCAGTGCTGTGATGATGCATTAATGCTG 11533
Db	64 CAATCAATGCTGTTATTCCTGCTGCTGTACTCGTGCCTGCTGCTGTTGATCTTATTCG 123
Qy	1154 TATTAGTGTGAAAATCAAGTCGGAATTTTAA 1187
Db	124 CTATCGTATCATCATCAATGATAATTAATAAA 157
RESULT 15	
LOCUS	COI42555 360 bp mRNA linear EST 17-JUN-2004
DEFINITION	ESR8J7226 Aspergillus flavus Normalized cDNA Expression Library
ACCESSION	Aspergillus flavus cDNA clone NARE095 5' end, mRNA sequence.
VERSION	COI42555
KEYWORDS	COI42555.1 GI:48896556
ORIGIN	EST.
ASPERGILLUS FLAVUS	
EUKARYOTA; FUNGI; ASCOMYCOTA; PEZIZOMYCOTINA; EUTROTIOMYCETES; EUROTIALES; TRICHOCOMACEAE; MITOSPORIC TRICHOCOMACEAE; ASPERGILLUS	
(Bases 1 to 360)	
Yu.J., Whitelaw,C.A., Niemman,W.C., Bhatnagar,D., Cleveland,T.E. Aspergillus flavus expressed sequence tags for identification of genes with putative roles in aflatoxin contamination of crops FEBS Microbiol. Lett. (2004) In press	
Contact: Yu J	
Food and Feed Safety Research Unit	
USA/ARS, Southern Regional Research Center	
1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA	
Tel.: 504 286 4405	
Fax: 504 286 4419	
Email: jinyu@errc.ars.usda.gov	
Contact Dr. Yu at USDA/ARS SRRC (jinyu@errc.ars.usda.gov) for clones	
JOURNAL COMMENT	

[illegible]

Location/Qualifiers
1. .412

```

/organism="Aspergillus flavus"
/mol_type="mRNA"
/strain="NRRL 3357"
/db_xref="taxon:5059"
/clone="NAFAU54"
/sex="asexual mycelia"
/cell_type="mycelia"
/dev_stage="developmental stages from 18 to 96 hours"
/lab_host="E. coli DH10B T1 resistant cells"
/clone_lib="Aspergillus flavus Normalized cDNA Expression
Library"
/note="Vector: pBluescript (SK+) (Stratagene), antibiotic
selection marker: Carbenicillin, Site_1: NotI, at the 5
prime end; Site_2: BclRI, at the 3 prime end; This
normalized cDNA_expression library was constructed using a
mixture of mycelial cells grown under eight different
medium conditions and harvested at 5 time points (18, 24,
48, 72, 96 hours). The poly-A sequence was trimmed off
before ligating to vector."

```

Query Match	3.8%;	Score 49.2;	DB 7;	Length 412
Best Local Similarity	58.0%;	Pred. No. 0.12;		
Matches	87;	Conservative	0;	Mismatches 23

	Sequences	53	Indels	0	Gaps	0
Oy	1028 TAACAAGAAATACAGAGTTCCAAATTGAACCTGATGTGCCAAACCAATTGTTGACCGTT					108
Db	108 TGACCCCGCAGCGGAGTCCCAATTGAGCCGCGGTACAGACAGAACTGCTTGATGCCG					167
Oy	1088 GTCAAGAAATTCCTGGTTGTGTGGTGTGTGTTCCAGATGTGTGTGATACATGCA					1147
Db	168 TGTACACCGGTACCGAGTGAATTGGTGTGTGTCTTCTGTGTGGGAGGTATGATGGA					227
Oy	1148 TAGCTGTATTAGTTGTGAAATTAAGTCG					1177
Db	228 TTGCTGTCTGATCCGCGATACACGAGAG					257

COL142023
100719

LOCUS	412 bp	mRNA	linear	EST 17-UTN-2004
DEFINITION	EST836564	<i>Aspergillus flavus</i>	Normalized cDNA	Expression library
ACCESSION	Aspergillus flavus	CDNA clone	MAFE081 5' end,	mRNA sequence.
RECORD	CO142023			

REFERENCE	AUTHORS	JOURNAL	COMMENT
1 (bases 1 to 412)	Yu, J., Whitelaw, C.A., Nierman, W.C., Bhatnagar, D. and Cleveland, T.E	Aspergillus flavus expressed sequence tags for identification of genes with putative roles in aflatoxin contamination of crops	FEMS Microbiol. Lett. (2004) In press
	Yu, J.		Contact: Yu, J

USDA/ARS, Southern Regional Research Center
1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA
Tel: 504 286 4405
Fax: 504 286 4419
Email: jinyu@srrc.ars.usda.gov
Contact Dr. Yu at USDA/ARS_SRRRC (jinyu@srcc.ars.usda.gov) for clone
information
PCR Primers
FORWARD: M13F
BACKWARD: M13R

Source

Seq primer: M13 Forward
POLYA=No.

source
1. .412

```

/mol type="RNA" 9...LIVAS-
/strain="NRRL 3357"
/db xref="Taxon:5059"
/clone="NAFCU81"
/sex="asexual mycelia"
/cell_type="mycelia"
/dev_stage="developmental stages from 18 to 96 hours"
/lab host="E. coli DH10B T1 resistant cells"
/clone_id="Aspergillus flavus Normalized cDNA Expression
library"
/note="vector: pBluescript (SK+) (Stratagene), antibiotic
selection marker: Carbenicillin; Site_1: NotI, at the 5
prime end; Site_2: EcoRI, at the 3 prime end; This
normalized cDNA_expression library was constructed using a
mixture of mycelial cells grown under eight different
medium conditions and harvested at 5 time points (18, 24,
48, 72, 96 hours). The poly-A sequence was trimmed off
before ligating to vector."
```

Query Match	3.8%;	Score 49.2;	DB 7;	Length 412;
Best Local Similarity	58.0%;	Pred. No. 0.12;		
Matches	87;	Conservative	0;	Mismatch

[illegible]

LOCUS

DEFINITION	EST823863 Aspergillus flavus Normalized cDNA Expression Library
ACCESSION	Aspergillus flavus cDNA Clone NAGDL94 5' end, mRNA sequence.
VERSION	COL48810
KEYWORDS	COL48810.1 GI:48902811
EST.	

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1
1 (bases 1 to 461)
Yu, J., Whittell, C.A., Nieman, W.C., Bhattacharya, D. and Cleveland, T.E.
Aspergillus clavus expressed sequence tags for identification of
genes with putative roles in aflatoxin contamination of crops
FEMS Microbiol. Lett. (2004) In press
Contact: Yu J

usda/ars, Southern Regional Research Center
1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA
Tel: 504 286 4405
Fax: 504 286 4419
Email: jinyu@srcc.ars.usda.gov
Contact Dr. Yu at USDA/ARS SRRC (jinyu@srcc.ars.usda.gov) for clone
information
PCR Primers
FORWARD: M13F
BACKWARD: M13R
Seq primer: M13 Forward
SOLYA=No.

/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="L.serriola"
/db_xref="taxon:4236"
/clone="QG5L17"
/lab_host="E.coli"
/clone_lib="QG_EFGHJ lettuce serriola"
/note="Vector: pBRCDNA5flab. The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformed into a custom medium-copy vector and size bias. Details of each source of RNA and library construction can be obtained at http://cspdb.ucdavis.edu/TAG_SEO=Not+found"

ORIGIN

Query Match 3.7%; Score 47.8; DB 5; Length 775;
Best Local Similarity 60.3%; Pred. No. 0.33;
Matches 79; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY 517 AATAAAGATTTTGGCAACGTTGCAACGATTTGATGCCCCAAAAGATA 576
DB 584 AATCTTATGTGTGATATTATAGCCCAATGGCACTGATTGTCAGAAAGTT 643
QY 577 GATCTGGGTTGATGTGCAACTGCAATTTATGCTGATTAATAGAGATTTCAG 636
DB 644 GGGAGTGGATTTATTTAGTTAGCTGTTATGAGAGTCATCGTTATGTAAGTTTCA 703
QY 637 CCAGCTTTGAT 647
DB 704 CCACCACTTAT 714

RESULT 21
LOCUS COL152593 447 bp mRNA linear EST 17-JUN-2004
DEFINITION EST827646 Aspergillus flavus Normalized cDNA Expression Library
ACCESSION COL152593
VERSION COL152593.1 GI:48906594
KEYWORDS EST.
SOURCE Aspergillus flavus
ORGANISM Aspergillus flavus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
1 (bases 1 to 447)
YU.J., Whitefaw,C.A., Niernan,W.C., Bhannagar,D. and Cleveland,T.E.
Aspergillus flavus expressed sequence tags for identification of
genes with putative roles in aflatoxin contamination of crops
FEMS Microbiol. Lett. (2004) In press
Contact: Yu J
Food and Feed Safety Research Unit
USDA/ARS, Southern Regional Research Center
1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA
Tel: 504 286 4405
Fax: 504 286 4419
Email: jinyu@arcc.ars.usda.gov
Contact Dr. Yu at USDA/ARS SRRC (jinyu@arcc.ars.usda.gov) for clone
information
PCR Primers
FORWARD: M13P
BACKWARD: M13R
Seq primer: M13 Forward
POLYA=No.

FEATURES

source 1..447
Location/Qualifiers
/organism="Aspergillus flavus"
/mol_type="mRNA"
/strain="NRRL 3357"
/db_xref="taxon:5059"

/clone="NA6FN24"
/sex="asexual mycelia"
/cell_type="mycelia"
/dev_stage="developmental stages from 18 to 96 hours"
/lab_host="E. coli DH10B T1 resistant cells"
/clone_lib="Aspergillus flavus Normalized cDNA Expression Library"
/note="Vector: pBluescript (SK+) (Stratagene), antibiotic selection marker; Cardenichillin; Site 1: NotI, at the 5 prime end; Site 2: EcoRI, at the 3 prime end; This normalized cDNA expression library was constructed using a mixture of mycelial cells grown under eight different medium conditions and harvested at 5 time points (18, 24, 48, 72, 96 hours). The poly-A sequence was trimmed off before ligating to vector."

ORIGIN

Query Match 3.7%; Score 47.6; DB 7; Length 447;
Best Local Similarity 57.3%; Pred. No. 0.32;
Matches 86; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
QY 1028 TAACCAAAAATTCAGAGTTTCAATTGAACCTGATGCCAACCGATTGTGACCGTT 1087
DB 143 TGACCCCGCAAGCGCGAGTCCCATTTAGCCCGGATACAGAACTGCTTATGCCG 202
QY 1088 GTCAAGAGATTCGCGTTGTGTGTTGTTGTTCCAGTCTGCTGATATACATGCA 1147
DB 203 TGTACGCCGTGACGAGATGATGTTGTTGTTCTGTGCGGAAAGTATATGCA 262
QY 1148 TAGCTGATTAAGTTGTTGAAAATCAAGTGG 1177
DB 263 TTGCTGTCTTGATCCCGATGACCAAGAG 292

RESULT 22
LOCUS COL137296 475 bp mRNA linear EST 17-JUN-2004
DEFINITION EST831967 Aspergillus flavus Normalized cDNA Expression Library
ACCESSION COL137296
VERSION COL137296.1 GI:48866274
KEYWORDS EST.
SOURCE Aspergillus flavus
ORGANISM Aspergillus flavus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
1 (bases 1 to 475)
YU.J., Whitefaw,C.A., Niernan,W.C., Bhannagar,D. and Cleveland,T.E.
Aspergillus flavus expressed sequence tags for identification of
genes with putative roles in aflatoxin contamination of crops
FEMS Microbiol. Lett. (2004) In press
Contact: Yu J
Food and Feed Safety Research Unit
USDA/ARS, Southern Regional Research Center
1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA
Tel: 504 286 4405
Fax: 504 286 4419
Email: jinyu@arcc.ars.usda.gov
Contact Dr. Yu at USDA/ARS SRRC (jinyu@arcc.ars.usda.gov) for clone
information
PCR Primers
FORWARD: M13P
BACKWARD: M13R
Seq primer: M13 Forward
POLYA=No.

FEATURES

source 1..475
Location/Qualifiers
/organism="Aspergillus flavus"
/mol_type="mRNA"
/strain="NRRL 3357"
/db_xref="taxon:5059"
/clone="NAFCP88"
/sex="asexual mycelia"

Db	675	DKATKAKKAKKAKKAKKAKKAKTATGKDRKAKAKAKAKAKATRKAKAKAMDTATTTATTAAD	734
Qy	694	ACAGATTGAAAAATTGATTGAAAGTAACGTGGNAATTCAAACATGAAGAATGTACATTA	753
Db	735	AAADKGAADKAKAKAKADAPDKRRWDAKDGRKAKAKAATAADADADADADAADAADA	794
Qy	754	CCATACCGAATCAAGTTATTAATGAGGTGAGCTCAAGGGTGGCTCAGAAAACCCAAATTG	813
Db	795	KADAAADDDRGDGGDKKKRKAQRKKKKKKDKAMGDKKAKKDKAAAKAKADAAGAKAKRR	854
Qy	814	GTATCAACGATGCTCCAAATGCAAAAAAGCAAGAAAGCTCTGTTGTATGAC	873
Db	855	RAGDKDKAKADAKAKAKAKKDDDAKATRAKATRAKKKKKKKKKKDKADAKA	914
Qy	874	CAGCTTAATAGTCCCAATTTACAGTTATG-AAAGAAATTTAGAGGAAATGCGTGAATAA	932
Db	915	KAKADKDDDDDKDADKADKKKKKORAKDKDDDDDKDKADKDKDKKDDADAK	974
Qy	933	CGACTCAGACCCAGAGACTATATTAAGAGTTGATCATTCGTGTGAGCCTTGACCTGT	992
Db	975	AWAAKAKDADADAKAKAKAKAADADADADAKADADAKAKAKADDAKAKAKADAVAA	1034
Qy	993	TGCGATTAAAGACATCAGAAAAAGGTTA	1020
Db	1035	DADAKAKAKRAKADADADAKAKAKA	1062

RESULT 27		
CNS0182P		
LOCUS	1101 bp	DNA linear GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC	
	BACN37D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.	

ACCESSION	AL108811
VERSION	AL108811.1
KEYWORDS	GI:5629115
SOURCE	GSS.
ORGANISM	<i>Drosophila melanogaster</i> (fruit fly)
	<i>Drosophila melanogaster</i>

ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
REFERENCE
AUTHORS

TITLE	Dietetic supervision
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006-EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a

<http://www.edgpb.ebi.ac.uk>. This *Drosophila* melanogaster BAC library (Dros BAC) was made by Alain Billard at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBel0BAC11.

FEATURES	SOURCE	Location/Qualifiers
ORIGIN		1..1101 /organism="Drosophila melanogaster" /mol_type="genomic DNA" /db_xref="taxon:7227" /clone="BACn37D10" /clone_1fb="DrosBAC" /plasmid="pBelobAct11" /note="end : SP6"

Query Match	3.6%	Score 45.6;	DB 9;	Length 1101;
Best Local Similarity	20.9%;	Pred. No. 0.75;		
Matches	99;	Conservative 162;	Indels	0;
			Gaps	0;
654	ACAGCTGAAAAATTTGATTCGAAGTACACCGCATTCACACGAAAGATGTCACATTA	753		
: : : : :	: : :			

Db 576 AAAAADPTKORNDARGAAAARAAAADAABAARFAAKTMAAAAAAAAAAAAAAAAAAAAA 635

Qy 754 CCATACCGAATCAAGTTATTATTAAGGTGACGCTCAAGGGTGGCTAGAAACACCCAAATTG 813

Db 636 GTTATTAAAAAAGGCKTGKGRAADODGGAAAAADMBRGRGGRBARBARAAADAAADAK 695

Qy 814 GTATACAGACTACTCCAAATGAAAAAGAAAAAGCCAGAGAAAGACTCTGTTGTATGAC 873

Db 696 RAAAAAAAAAAAAAAAAAADAGRKKWKDGDKGGRATKTAAMAKGRKMDGTATAMWTW 755

Qy 874 CAGCTTAATAGTGCCTATTTCAGTTTATGTAAGGAATTTGAGGAAATGCGTGAATAATAC 933

Db 756 DATWADPTKATTDIDAKAAAAGRRKRPARKTARBGRRARTRRRAAWAAGGRABAGARRA 815

Qy 934 GACTCAGAACCCAGAGACTTATTTATTAAGAGTTAATCATTTCTGTGACCTTTGACGT 993

Db 816 RAARBAADDBRDWDMWMAAAAAAAAAAAAMTTWRBWMWDWMDTRWDDTTAAWMDPARA 875

Qy 994 GCGATTAAAGAACATCAGAAAAGGTTACAAGCATTAACACAAAATAAGAGGTTCCAAAT 1053

Db 876 RARRRRRRRRRRARRARAADDTYTKRMAADTTDKTTKMTTDDDDWDPAKRW 935

Qy 1054 GAACCTGATGTCAAACCCAGTTGTGTGGAACCGTTGTAAGAGATTCCGTTGTGTTGGT 1113

Db 936 WAAARADAGWKRRDARADWAAATADDDGWDKMWGGRGKRGSKDKKRWDKGTGKKDDDW 995

Qy 1114 GGTGTGGTTCAAGGTCGTGAGATACAGATCAATAGCTGATTAAGTTGGA 1166

Db 996 DKTTTRDMMWMTTKMDWMDDBRGGRGRWTRKRGWAMWRADAWARDDTDSKD 1048

RESULT	28				
B0866497					
LOCUS					
DEFINITION		716 bp	mRNA	linear	EST 14-AUG-2002
	B0866497	OGC8C07.Y9.ab1	OG ABCDI lettuce salinas	Lactuca sativa	cDNA clone
	OGC8C07		mRNA sequence.		
ACCESSION					
VERSION					
KEYWORDS					
	B0866497.1	GI:22251962			
	EST.				

KEYWORDS	SOURCE	ORGANISM
EST.	<i>Lactuca sativa</i>	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Asterales; Asteraceae; Cichorioideae;

REFERENCE
AUTHORS

Cichowski, J.,
1 (bases 1 to 716)
Kozik, A., Michemore, R. W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lih, H., van Damme, W., Lavelle, D., Chevalier, P., Ziegler, J.,
Ellison, P., Kolman, J., Slabaugh, M. S., Livingston, K., Zhou, Y.,

TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project
JOURNAL <http://compgenomics.ucdavis.edu/>
COMMENT Unpublished (2002)
Contact: Alexander Kozik [R.W.Mitchellmore]
Department of Vegetable Crops, R.W.Mitchellmore Lab
University of California at Davis (UCD)

Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akoxil@atgc.org [michelemorevegmail.ucdavis.edu]
belongs to contig_Q3_CA_contig6356, see <http://cgpdb.ucdavis.edu/>
for details
plate: QGC8. row: C column: 07.
Location/Qualifiers
1. .716
/organism="Lactuca sativa"

```

/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="QGC8C07"
/lab_host="E. coli"
/clone_lib="QC_ARCD1_lentuce_salinas"
/notes="Vector: pBRCDNAStab; The library was constructed

```

Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformation made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpb.udavis.edu/TNG_T150SE-Chemical_induction

..._oey=1G1AGCUCGGC

Query Match	3.6%;	Score	46.2;	DB	5;	Length	745
Best Local Similarity	59.5%;	Pred	Wa	0	0		

Actual	Pred.
59.5%	59.5%

A T A M M A S S E - C O N F I D E N T I A L

|||||

.....GCTTGGTTATGAGC

5GTTTGATGTTGCAACTGC

5A TTTGATGTTAGTTTCAAGC

MGAT 647

—

1101
melanogaster genome

OL PROBAB Library

1
2
3
4
5

01407001

melanogaster (fruit)

Metazoa; Arthropoda;

Drosophila

CO 1101)

mission

6 EYBY codon

.genoscope.cns.fr)

on with the Eurycea

edg. BAC) was 100%.

Polymorphisme Humain

re Payan. It has been

2024/05/07: 11:21

.1101

type="genomic" data-bbox="100 100 115 250"/>

```

_Xref="faxon:7227"

```

one_lib="DrosBAC"

```
end pre10BAC11"
end"end : end"
```

3.6%; Score 46

21.06; Pled. No.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

0973

Query Match	Best Local Similarity	Score	DB	Length
Matches 777	Conservative 0	Mismatches 54	Indels 0	Gaps 0
QY	517	AATPAAGATATTTTGACACAGTTGCACAGATTGCACATTGTTATGCCCAAAAAAGATA	576	
Db	562	AACTTGAATGTGTGCATATTTATACCCCAANTGGCACACTGTAATTCCTCAAGAAAGATT	621	
QY	577	GGATCGGGTTTGAATGCTTGCACACTGCACATTTAATGCTGATTTATATAGAAGATTTCAG	636	
Db	622	GGGAGTGATTTGAATGTTAGTTGACGTGTTATGGAAGCATCGTTATGTAAAGATTTC	681	
QY	637	CCAGCTTTGAT	647	
Db	682	CCACAAGTTAT	692	

RESULT 34					
BM165227/c					
LOCUS	BM165227	697 bp	mRNA	linear	EST 04-DEC-2001
DEFINITION	BM165227				
	EST567750 PyBS plasmidium yoeI:: yoeII cDNA clone pYCMJ47 5' end,				
ACCESSION	BM165227				
VERSION	BM165227.1	GI:17310908			
KEYWORDS	Est.				
SOURCE	Plasmidium yoeI:: yoeII::				

REFERENCE
1. (baes 1 to 697)
Carlton, J. M., Daly, T. M., Long, C. A., Bergman, L. W., Valdivia, A. B., Frazer, C. M. and Carucci, D. J.
Plasmidom voelji; EST project of the

11715
JOURNAL
COMMENT
Plasmodium yoelii EST project at TIGR
Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9339
Fax: 301-838-0335

FEATURES
source
for clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADP.
Location/Qualifiers
1.697
/organism="Plasmodium yoelii yoelii"
/mol_type="mRNA"
/strain="17XL"
/sub_species="yoelii"

collected from BALB/cByJ mice 20-25% parasitemia, blood was separated into plasma and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and rRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, unreacted termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to

HybridZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HybridZAP vector and plasmid DNA isolated."

ORIGIN

Query Match 3.5%; Score 45; DB 4; Length 697;
 Best Local Similarity 47.4%; Pred. No. 1.7;
 Matches 135; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

Db 1014 AGGTTACAGCATTAAACAAAATCAGAGTTCATTCAGTCCAAACCA 1073
 648 AGGATCTAGATTAATAAAGTATCTTACCAGAGCTAAAGCAGAACATA 589

Qy 1074 GTTGTGACCGTGTCAAGAGATTCCTGTTGTGGTGGTGTTCAGGTCTGG 1133
 Db 588 AGTATATGAGTAGGTGATGAAGAGTGGAGGTGTGATTTAGAAAAAGAGTAG 529

Qy 1134 TGGATACAGTCAATGCTGTATTAGTTGGAATCAAGTGGAAATTTTACAGAA 1193
 Db 528 TGATGAAGAAGTGGAGGTGTGATTTAGAAAAAGAGTGAAGAAAGTGG 469

Qy 1194 AACTTTGAAAATCCAGATTATTTTCATATATGTTTACTGGGTTGATTGGAAGCAAC 1253
 Db 468 AGGTGTGAATTTTATAGATTAATGAAGTGTGAAGAGTGAAGGTGTTGAAGTTT 409

Qy 1254 AGAAGTGTACTTGAAGAAAACAGAACTATATAGGTTATA 1298
 Db 408 AGAAGAAGATGATGATGAAGATGATGATGATGATGAAGATTTAGA 364

RESULT 35

BM169836 780 bp mRNA linear EST 04-DEC-2001
 LOCUS EST572359 PyBS Plasmodium yoelii yoelii cDNA clone PYCPX37 5' end,
 DEFINITION mRNA sequence.
 ACCESSION BM169836
 VERSION BM169836.1 GI:17303068
 KEYWORDS EST.
 SOURCE Plasmodium yoelii yoelii
 ORGANISM Plasmodium yoelii yoelii
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 780)
 AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
 Fraser,C.M. and Carucci,D.J.
 TITLE Plasmodium yoelii EST project at TIGR
 JOURNAL Unpublished (2001)
 COMMENT Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADF.

FEATURES

source 1..780
 Location/Qualifiers
 /organism="Plasmodium yoelii yoelii"
 /mol_type="mRNA"
 /strain="17XL"
 /sub_species="yoelii"
 /db_xref="taxon:73239"
 /clone="PYCPX37"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /clone_lib="PyBS"
 /note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with FY11XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was

isolated using the guanidium isothiocyanate method, and mRNA isolated using Oligo(dt)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybridZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HybridZAP vector and plasmid DNA isolated."

ORIGIN

Query Match 3.5%; Score 45; DB 4; Length 780;
 Best Local Similarity 47.4%; Pred. No. 1.8;
 Matches 135; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

Db 1014 AGGTTACAGCATTAAACAAAATCAGAGTTCATTCAGTCCAAACCA 1073
 608 AGGATCTAGATTAATAAAGTATCTTACCAGAGCTAAAGCAGAACATA 549

Qy 1074 GTTGTGACCGTGTCAAGAGATTCCTGTTGTGGTGGTGTTCAGGTCTGG 1133
 Db 548 AGTATATGAGTAGGTGATGAAGAGTGGAGGTGTGATTTAGAAAAAGAGTAG 489

Qy 1134 TGGATACAGTCAATGCTGTATTAGTTGGAATCAAGTGGAAATTTTACAGAA 1193
 Db 488 TGATGAAGAAGTGGAGGTGTGATTTAGAAAAAGAGTGAAGAAAGTGG 429

Qy 1194 AACTTTGAAAATCCAGATTATTTTCATATATGTTTACTGGGTTGATTGGAAGCAAC 1253
 Db 428 AGGTGTGAATTTTATAGATTAATGAAGTGTGAAGAGTGTGTAAGTTT 369

Qy 1254 AGAAGTGTACTTGAAGAAAACAGAACTATATAGGTTATA 1298
 Db 368 AGAAGAAGATGATGATGAAGATGATGATGATGATGAAGATTTAGA 324

RESULT 36

BM161967 805 bp mRNA linear EST 04-DEC-2001
 LOCUS EST564490 PyBS Plasmodium yoelii yoelii cDNA clone PYCKL43 5' end,
 DEFINITION mRNA sequence.
 ACCESSION BM161967
 VERSION BM161967.1 GI:17307648
 KEYWORDS EST.
 SOURCE Plasmodium yoelii yoelii
 ORGANISM Plasmodium yoelii yoelii
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 805)
 AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
 Fraser,C.M. and Carucci,D.J.
 TITLE Plasmodium yoelii EST project at TIGR
 JOURNAL Unpublished (2001)
 COMMENT Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADF.

FEATURES

source 1..805
 Location/Qualifiers
 /organism="Plasmodium yoelii yoelii"
 /mol_type="mRNA"
 /strain="17XL"
 /sub_species="yoelii"

DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL069440

VERSION AL069440.1 GI:4949583

KEYWORDS GSS.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

REFERENCE Genoscope.

AUTHORS Direct Submission

TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

JOURNAL

COMMENT

- Web : www.genoscope.cns.fr

- Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Location/Qualifiers

1..1101

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/db_xref="taxon:7227"

/clone="BACR29P01"

/clone_11b="RPCI-98"

/note="end : TET3"

ORIGIN

Query Match 3.4%; Score 44.6; DB 9; Length 1101;

Best Local Similarity 33.7%; Pred. No. 2.5;

Matches 114; Conservative 54; Mismatches 170; Indels 0; Gaps 0;

QY 480 TTTATTATCCCATTTTATCCCAATGTTATCAGTACGATAAAGATTTTGCACAACGT 539

DB 742 TTTATTAAWMTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 801

QY 540 TGCACAGATTGCACATGTTATGCCCAAAAAGATGATCTGGGTTGATGTTGCAAC 599

DB 802 WKATATATWTAATWTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 861

QY 600 TGAATTTATGCTGATGTTATAGAGATTTCAACGCTTGTATTAATGACGTGT 659

DB 862 TTTTMMWTAATWTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 921

QY 660 TCAGGTTCTAGAAAGTATCTGGAAGTCCCAAGAGTTGAAAAAATTGATGGAAG 719

DB 922 TTTTATTTTAAATTTGATTAATAAAAWATTAATTAATTAATTAATTAATTAATTA 981

QY 720 TAACGGGAATTCAACATGAAGATGATACATACGAAATCAAGTTATTAATGAGG 779

DB 982 ATATWGRATTTKAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1041

QY 780 TGAAGTCAAGGCTGCTCAAAACCAACCAATTGGTAT 817

DB 1042 GKGKGKKTGKGKGKGTTGGGDAAAARAAKAKGKT 1079

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 26, 2005, 03:01:13 ; Search time 483 Seconds
(without alignments)
14118.008 Million cell updates/sec

Title: US-10-069-062-6
Perfect score: 1299
Sequence: 1 atgcacaaagcatttcgtc.....aagactataggtttataa 1299

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*

- 1: geneseqn19808:*
- 2: geneseqn19908:*
- 3: geneseqn20008:*
- 4: geneseqn20018:*
- 5: geneseqn20018:*
- 6: geneseqn20028:*
- 7: geneseqn20028:*
- 8: geneseqn20038:*
- 9: geneseqn20038:*
- 10: geneseqn20038:*
- 11: geneseqn20038:*
- 12: geneseqn20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1299	100.0	1299	4 AAD02792	Aad02792 Candida a
2	1299	100.0	1763	4 AAD02791	Aad02791 Candida a
3	1286.2	99.0	1299	4 AAS23422	Aas23422 Candida a
4	1286.2	99.0	1299	6 ABZ31736	Abz31736 Candida a
5	521	40.1	577	4 AAD02788	Aad02788 S. cerevi
6	470.4	36.2	547	4 AAD02787	Aad02787 S. cerevi
7	132	10.2	1356	6 AAL40805	Aal40805 Nucleic a
8	132	10.2	1356	6 AAD31013	Aad31013 Yeast pho
9	132	10.2	1356	6 ABK96801	Abk96801 S. cerevi
10	132	10.2	1356	12 ADF23287	Adf23287 MevB oper
11	132	10.2	1356	12 ADI82315	Adi82315 S. cerevi
12	132	10.2	1356	12 ADI14834	Adi14834 Yeast pho
13	132	10.2	4482	12 ADF23291	Adf23291 Mevalonat
14	132	10.2	4482	12 ADI82319	Adi82319 Mevalonat
15	132	10.2	5051	12 ADF23294	Adf23294 Recombina
16	132	10.2	5051	12 ADI82322	Adi82322 Amorphin-4
17	132	10.2	5963	12 ADF23295	Adf23295 Recombina
18	132	10.2	5963	12 ADI82323	Adi82323 Amorphin-4
19	132	10.2	7681	6 AAD31026	Aad31026 Operon D
20	132	10.2	7683	6 AAD31023	Aad31023 Operon A
21	132	10.2	7695	6 AAD31024	Aad31024 Operon B

22	132	10.2	8224	6 AAD31027	Aad31027 Operon E
23	132	10.2	8235	6 AAD31025	Aad31025 Operon C
24	132	10.2	8400	6 AAD31029	Aad31029 Operon G
25	132	10.2	9253	12 ADF23289	Adf23289 Isopenic
26	132	10.2	9253	12 ADI82317	Adi82317 Amorphin-4
27	132	10.2	13917	6 AAD31037	Aad31037 Placid c
28	132	10.2	14623	6 AAD31039	Aad31039 Placid c
29	132	10.2	14623	6 AAD31041	Aad31041 Placid c
30	67.6	5.2	1506	10 ADB69813	ADB69813 C. neofo
31	66	5.1	1356	8 ABT19226	ABT19226 Aspergill
32	65	5.0	65	6 ABZ28940	ABZ28940 Candida g
33	65	5.0	90	4 AAS23667	Aas23667 Tetraacycl
34	62.6	4.8	1455	8 ABT21056	ABT21056 Aspergill
35	62.6	4.8	1509	8 ABT20458	ABT20458 Aspergill
36	62.6	4.8	1509	8 ABT18642	ABT18642 Aspergill
37	62.6	4.8	3508	8 ABT18048	ABT18048 Aspergill
38	62.6	4.8	3509	8 ABT19862	ABT19862 Aspergill
39	53.2	4.1	869	8 ABZ51452	ABZ51452 Aspergill
40	48.2	3.7	2000	8 ADA71938	Ada71938 Rice gene
41	46.8	3.6	586	3 AAF08722	AAF08722 Fusarium
42	45	3.5	2000	8 ADA71938	Ada71938 Rice gene
43	44.6	3.4	1356	12 ADL03441	Adl03441 DNA encod
44	43	3.3	1341	8 ACA39227	ACA39227 Prokaryot
45	43	3.3	119211	4 AAF28553	AAF28553 Genomic F

ALIGNMENTS

RESULT 1	AAD02792	standard; DNA; 1299 BP.
ID	AAD02792	
XX		
AC	AAD02792;	
XX		
DT	31-MAY-2001	(first entry)
XX		
DE	Candida albicans phosphomevalonate kinase (PMK) coding sequence.	
XX		
KW	Phosphomevalonate kinase; PMK; ERG8; anti-fungal agent; diagnosis;	
XX		
OS	Candida albicans.	
XX		
FM	Key	Location/Qualifiers
FT	CDS	1..1299
FT	FT	/*tag= a
FT	FT	/product= "C. albicans phosphomevalonate kinase (PMK)"
XX		/transl_except= (pos:730..732, aa:Glu)
PN	MO200114533-A2.	
XX		
PD	01-MAR-2001.	
XX		
PF	15-AUG-2000; 2000WO-GB003100.	
XX		
PR	21-AUG-1999; 99GB-00019766.	
XX		
PA	(ASTR) ASTRAZEMECA AB.	
XX	(ASTR) ASTRAZEMECA UK LTD.	
XX		
PI	Rosamond JDC, Schnell NF;	
XX		
DR	WPI, 2001-218441/22.	
XX		
DR	P-PSDB; AAY72679.	
XX		
PT	New polypeptides and polynucleotides (ERG8) from Candida albicans, useful	
XX	in assays for identifying inhibitors of phosphomevalonate kinase activity	
XX	and as reagents for diagnosing C. albicans infection.	
PS	Claim 6; Page 26; 29pp; English.	
XX		
XX	The present sequence is phosphomevalonate kinase (PMK; ERG8) coding	

CC sequence from *Candida albicans*. The ERG8 protein is useful in an assay
 CC for identifying compounds that inhibit phosphomevalonate kinase (PMK)
 CC activity. These inhibitors are useful as anti-fungal agents. The ERG8 DNA
 CC and protein are also useful as reagents for diagnosing *C. albicans*
 CC infection

XX
 SQ Sequence 1299 BP; 439 A; 220 C; 273 G; 367 T; 0 U; 0 Other;

Query Match 100.0%; Score 1299; DB 4; Length 1299;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGTCAAAAGCATTAGTGCACCTGGAAGAAAGCATTTCTGCTGGTGAATATTTGGTTCTT 60
DB 1 ATGTCAAAAGCATTAGTGCACCTGGAAGAAAGCATTTCTGCTGGTGAATATTTGGTTCTT 60
QY 61 GAGCCAAATTTATGATGCTATGAGACAGATTGTCATCAAGATGATGATGATTTTAA 120
DB 61 GAGCCAAATTTATGATGCTATGAGACAGATTGTCATCAAGATGATGATGATTTTAA 120
QY 121 CCAAAAGAACCGATTGGAAGAAATCTAGATCAAAATTTCTTCAACCCCAATTTGCAAC 180
DB 121 CCAAAAGAACCGATTGGAAGAAATCTAGATCAAAATTTCTTCAACCCCAATTTGCAAC 180
QY 181 GGAGAAATGGGAATATATCATATCATCAAAATACAGAGAACCCAGAGAAATTCAGTCACG 240
DB 181 GGAGAAATGGGAATATATCATATCATCAAAATACAGAGAACCCAGAGAAATTCAGTCACG 240
QY 241 ATAAATCCATTTTATGAGCACTATATCATGCTTTTATGCTTATATTTCAACCCGAA 300
DB 241 ATAAATCCATTTTATGAGCACTATATCATGCTTTTATGCTTATATTTCAACCCGAA 300
QY 301 GCATTTGATCTTGAATATCATATTTACTCAGACCCCGATATCATTTCAAGAAATATCT 360
DB 301 GCATTTGATCTTGAATATCATATTTACTCAGACCCCGATATCATTTCAAGAAATATCT 360
QY 361 GAAACCAAGACATCTCCGAATGAGAAAAAATTTCTTTACCATTTCTGTCCTATACC 420
DB 361 GAAACCAAGACATCTCCGAATGAGAAAAAATTTCTTTACCATTTCTGTCCTATACC 420
QY 421 GAAGTGAAGAAACCGGATTAGTTCATGCGAGGATTAGTGTCAAGTTGGCCCAAGT 480
DB 421 GAAGTGAAGAAACCGGATTAGTTCATGCGAGGATTAGTGTCAAGTTGGCCCAAGT 480
QY 481 TTATATCCATTTTATCCCAATGTTATCAGTACGAAATAAAGATATTTTGCACAACGT 540
DB 481 TTATATCCATTTTATCCCAATGTTATCAGTACGAAATAAAGATATTTTGCACAACGT 540
QY 541 GCACGATTTGCATTTGATGCCCCCAAAAAGATAGATGCGGTGGTTGATTTGCAACT 600
DB 541 GCACGATTTGCATTTGATGCCCCCAAAAAGATAGATGCGGTGGTTGATTTGCAACT 600
QY 601 GCAATTTATGCTCTGATTTGTATATAGAGATTTTCAAGCAGCTTTGATTAATGAGTGT 660
DB 601 GCAATTTATGCTCTGATTTGTATATAGAGATTTTCAAGCAGCTTTGATTAATGAGTGT 660
QY 661 CAGGTTCTAGAAAGTGATCTCGAAGAGTTCCCAAGAGTTGAAAAAATTTGATGAAAGT 720
DB 661 CAGGTTCTAGAAAGTGATCTCGAAGAGTTCCCAAGAGTTGAAAAAATTTGATGAAAGT 720
QY 721 AACTGGGAATTCAAACATGAAAGATGATACATACATACGGAATTAAGTTATATGAGTGT 780
DB 721 AACTGGGAATTCAAACATGAAAGATGATACATACATACGGAATTAAGTTATATGAGTGT 780
QY 781 GACGTCAGAGGTGCTCAGAAACACCAATTTGATACAGAGTCTCAATGGAAGAAAG 840
DB 781 GACGTCAGAGGTGCTCAGAAACACCAATTTGATACAGAGTCTCAATGGAAGAAAG 840
QY 841 GAAAGCCAGAAAGAAAGCTCTGTTGTATGACAGCTTAATAGTGCCTTAATTTACAGTTT 900
DB 841 GAAAGCCAGAAAGAAAGCTCTGTTGTATGACAGCTTAATAGTGCCTTAATTTACAGTTT 900
QY 901 ATGAAGGAATTAGAGGAATGCTGTAAGAAATATCAACTACAGCCAGAGACTTATATTTAA 960

```

```

DB 901 ATGAAGGAATTAGAGGAATGCTGTAAGAAATATCAACTACAGCCAGAGACTTATATTTAA 960
QY 961 GAGTTAATCATTTCTGTTAGCCTTTGACTGTGCGATTAAGAAATCAAGAAAGGGTTA 1020
DB 961 GAGTTAATCATTTCTGTTAGCCTTTGACTGTGCGATTAAGAAATCAAGAAAGGGTTA 1020
QY 1021 CAAGCATTAACAAAAAATCAGAGTTCATTTGAACTGTAATGTCCAAACCCCAATTTGTTG 1080
DB 1021 CAAGCATTAACAAAAAATCAGAGTTCATTTGAACTGTAATGTCCAAACCCCAATTTGTTG 1080
QY 1081 GACCGTTGCAAGATTCCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1140
DB 1081 GACCGTTGCAAGATTCCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1140
QY 1141 GATGCAATAGCTGTATTTAGTTGTTGAAATCAAGTGGGAAATTTTAAAGCAGAAATCTCTT 1200
DB 1141 GATGCAATAGCTGTATTTAGTTGTTGAAATCAAGTGGGAAATTTTAAAGCAGAAATCTCTT 1200
QY 1201 GAAATTCAGATTTATTTTATATGTTTACTGCGGTGATTTTGAAGCAACAGAGAGT 1260
DB 1201 GAAATTCAGATTTATTTTATATGTTTACTGCGGTGATTTTGAAGCAACAGAGAGT 1260
QY 1261 GTACTTGAAGAAAAACAGAGACTATATAGTTTATTA 1299
DB 1261 GTACTTGAAGAAAAACAGAGACTATATAGTTTATTA 1299

```

RESULT 2
 AAD02791
 ID AAD02791 standard; DNA; 1763 BP.

XX AAD02791;
 AC
 XX
 DT 31-MAY-2001 (first entry)

DE *Candida albicans* phosphomevalonate kinase (PMK) gene (ERG8 gene).
 XX
 KW Phosphomevalonate kinase; PMK; ERG8; anti-fungal agent; diagnosis;
 infection; de.

OS *Candida albicans*.

XX
 FH Location/Qualifiers

FT 89..1387
 FT CDS
 FT /tag= a
 FT /product= "C. albicans phosphomevalonate kinase (PMK)"
 FT /transl_except= (pos:818..820, aa:Glu)

XX MO200114533-A2.

XX 01-MAR-2001.

XX 15-AUG-2000; 2000WC-GB003100.

XX 21-AUG-1999; 99GB-00019766.

XX (ASTR) ASTRAZENECA AB.
 PA (ASTR) ASTRAZENECA UK LTD.

XX Rosemond JDC, Schnell NF;

XX WPI; 2001-218441/22.

XX P-PSDB; AAY72679.

XX New polypeptides and polynucleotides (ERG8) from *Candida albicans*, useful
 PT in assays for identifying inhibitors of phosphomevalonate kinase activity
 PT and as reagents for diagnosing *C. albicans* infection.

XX Claim 6; Fig 1; 29pp; English.

XX The present sequence is phosphomevalonate kinase (PMK) gene (ERG8 gene)
 CC from *Candida albicans*. The ERG8 protein is useful in an assay for

CC Identifying compounds that inhibit phosphomevalonate kinase (PMK)
 CC activity. These inhibitors are useful as anti-fungal agents. The ERG8 DNA
 CC and protein are also useful as reagents for diagnosing *C. albicans*
 CC infection
 CC

Sequence 1763 BP; 582 A; 300 C; 349 G; 530 T; 0 U; 2 Other;

Query Match 100.0%; Score 1299; DB 4; Length 1763;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGTCAAAAGCATTTAGTCACCTGGAAAAGCATTTCTGCTGGTGGATTTGGTTCTT 60
DB 89 ATGTCAAAAGCATTTAGTCACCTGGAAAAGCATTTCTGCTGGTGGATTTGGTTCTT 148
QY 61 GAGCCAAATTATGATGCTTATGTCAGAGTTGTATCAAGATGATGATGATTTATACA 120
DB 149 GAGCCAAATTATGATGCTTATGTCAGAGTTGTATCAAGATGATGATGATTTATACA 208
QY 121 CCAAAAGAACGATTTGAAAGATCTAGATCAAAATTTCTTCAACCCCAATTTGCAAC 180
DB 209 CCAAAAGAACGATTTGAAAGATCTAGATCAAAATTTCTTCAACCCCAATTTGCAAC 268
QY 181 GGAGAAATGGAAATATCAATATCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
DB 269 GGAGAAATGGAAATATCAATATCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 328
QY 241 ATAAATCCATTTTGAAGCACTATATCATGCTTTTGTATGCTTATATTCACCGACCGAA 300
DB 329 ATAAATCCATTTTGAAGCACTATATCATGCTTTTGTATGCTTATATTCACCGACCGAA 388
QY 301 GCATTGATCTTGAATCATCATCTTCTGACACCTGATATCATTCACAGAGATGATCT 360
DB 389 GCATTGATCTTGAATCATCATCTTCTGACACCTGATATCATTCACAGAGATGATCT 448
QY 361 GAAACCAAGACATCTCGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 449 GAAACCAAGACATCTCGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 508
QY 421 GAAATGGAAGAAACCGGATTTAGTTCATGGCAGAGATTTGATGATGATTTGGCAAGT 480
DB 509 GAAATGGAAGAAACCGGATTTAGTTCATGGCAGAGATTTGATGATGATTTGGCAAGT 568
QY 481 TTAATATCCATTTTATCCCAATGTTATCAATGATTAAGATATTTTGGCAAGT 540
DB 569 TTAATATCCATTTTATCCCAATGTTATCAATGATTAAGATATTTTGGCAAGT 628
QY 541 GCACAGATTTGACATTTGTTATGCCCCAAAAGATAGATCTGGGTTTATGTTGCAACT 600
DB 629 GCACAGATTTGACATTTGTTATGCCCCAAAAGATAGATCTGGGTTTATGTTGCAACT 688
QY 601 GCAATTTATGCTGATTTGTTATAGAAATTTGACCGAGCTTTGATTAATGACGTTT 660
DB 689 GCAATTTATGCTGATTTGTTATAGAAATTTGACCGAGCTTTGATTAATGACGTTT 748
QY 661 CAGGTTCTGAAAGTATCTGAGAGAGTTCCCAAGAGTTGAAAAATTTGATTAAGT 720
DB 749 CAGGTTCTGAAAGTATCTGAGAGAGTTCCCAAGAGTTGAAAAATTTGATTAAGT 808
QY 721 AACTGGGAATTCAAACATGAAAGATGTACATTAACATGGAATCAAGTTATTAATGGGT 780
DB 809 AACTGGGAATTCAAACATGAAAGATGTACATTAACATGGAATCAAGTTATTAATGGGT 868
QY 781 GACGTCAAGGTGGCTCAGAAAACCAAAATTTGATTCAGAGATCTCCAAATGAAAAAG 840
DB 869 GACGTCAAGGTGGCTCAGAAAACCAAAATTTGATTCAGAGATCTCCAAATGAAAAAG 928
QY 841 GAAAGCCAGAAAGAAAGCTCTGTTGTTATGACAGATCTTAATGATCAAGTTT 900
DB 929 GAAAGCCAGAAAGAAAGCTCTGTTGTTATGACAGATCTTAATGATCAAGTTT 968
QY 901 ATGAAGAAATTTGAGGAAATGCGTGAATAATGACTCAGACCCAGAGACTTATTAATAA 960

```

```

DB 989 ATGAAGAAATTTGAGGAAATGCGTGAATAATGACTCAGACCCAGAGACTTATTAATAA 1048
QY 961 GAGTTAGATCATCTTGTGAGGCTTTGACTGTGGATTAAGACATCAGAAAAGGTTA 1020
DB 1049 GAGTTAGATCATCTTGTGAGGCTTTGACTGTGGATTAAGACATCAGAAAAGGTTA 1108
QY 1021 CAGCATTTACCAAAAATCAGAGGTTCCAAATTTGAACCTGATGTCAAAACCAAGTTGTTG 1080
DB 1109 CAGCATTTACCAAAAATCAGAGGTTCCAAATTTGAACCTGATGTCAAAACCAAGTTGTTG 1168
QY 1081 GACCGTTGCAAGAAATCTGCTGTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1140
DB 1169 GACCGTTGCAAGAAATCTGCTGTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1228
QY 1141 GATGCAATAGCTGTATTAATGTTGAAAAATCAAGTGGGAAATTTTAAAGCAGAAAATCTTT 1200
DB 1229 GATGCAATAGCTGTATTAATGTTGAAAAATCAAGTGGGAAATTTTAAAGCAGAAAATCTTT 1288
QY 1201 GAAAAATCCAGATTTATTTCAATATGTTTATCTGGTTGATTTGAAAGAGCAACAGAGGT 1260
DB 1289 GAAAAATCCAGATTTATTTCAATATGTTTATCTGGTTGATTTGAAAGAGCAACAGAGGT 1348
QY 1261 GTACTTGAAGAAAAACCGAAGACTATATAGTTTATTA 1299
DB 1349 GTACTTGAAGAAAAACCGAAGACTATATAGTTTATTA 1387

```

RESULT 3
 AAS23422
 ID AAS23422 brandard; DNA; 1299 BP.
 XX AAS23422;
 AC
 XX
 DT 04-DEC-2001 (first entry)
 XX
 DE Candida albicans essential gene CaYMR220W (ERG8).
 XX
 KW Gene identification; essential gene; GRACE; pathogenic fungus;
 KM gene replacement and conditional expression; fungal infection; da.
 XX
 OS Candida albicans.
 XX
 PN WO200160975-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US005551.
 XX
 PR 18-FEB-2000; 2000US-0183534P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Roemer T, Jiang B, Boone C, Bussey H;
 XX
 DR WPI; 2001-489080/53.
 XX
 PT P-PSDB; AAU15094.
 XX
 PT Identifying genes essential to fungal metabolisms and identifying
 PT potential therapeutic agents that target these genes.
 XX
 PS Claim 22; Page 162; 324pp; English.
 XX
 CC The present invention relates to novel methods for constructing fungal
 CC strains useful for identification and validation of gene products as
 CC targets for therapeutic agents, for creating a collection of identified
 CC essential genes, and screening assays for the discovery of new drugs. The
 CC invention provides the GRACE (gene replacement and conditional
 CC expression) method for the construction of mutant organisms referred to
 CC as GRACE strains of the organism. The invention can be applied to any
 CC organism, particularly a pathogenic fungus e.g. *Candida albicans*.
 CC *Aspergillus fumigatus* and *Cryptococcus neoformans*. The methods are useful
 CC to identify agents that may be used in the treatment of fungal
 CC infections. AAS23381-AAS23442 represent *C. albicans* essential genes


```

XX Sequence 1299 BP; 439 A; 221 C; 271 G; 368 T; 0 U; 0 Other;
SQ
Query Match          99.0%; Score 1286.2; DB 4; Length 1299;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1291; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATGTCAAAAGCATTGTAGTGCACCTGGAAAAGCATTTCTGCTGTGTGATATTTGGTTCTT 60
DB 1 ATGTCAAAAGCATTGTAGTGCACCTGGAAAAGCATTTCTGCTGTGTGATATTTGGTTCTT 60
QY 61 GAGCCAAATTTATATGATGTTATGTGACAGCATTCGTCAATCAAGATGATGAGTTATPACA 120
DB 61 GAGCCAAATTTATATGATGTTATGTGACAGCATTCGTCAATCAAGATGATGAGTTATPACA 120
QY 121 CCAAAAAGAACGATTTGAAAAGATCTAGAAATCAAAATTTCTTACCCTCAATTTGGAAC 180
DB 121 CCAAAAAGAACGATTTGAAAAGATCTAGAAATCAAAATTTCTTACCCTCAATTTGGAAC 180
QY 181 GAGAAATGGGAATATCAATATATCATCAAAATACAGAAAGCCAGAAAGTTGAGTCAAGC 240
DB 181 GAGAAATGGGAATATCAATATATCATCAAAATACAGAAAGCCAGAAAGTTGAGTCAAGC 240
QY 241 ATAAATCCATTTTATAGGCAATATATTCATGTTTACGTTATATTTCAACCGACGAA 300
DB 241 ATAAATCCATTTTATAGGCAATATATTCATGTTTACGTTATATTTCAACCGACGAA 300
QY 301 GCATTGATCTTGAATATCATATTTTACTGACAGCCCTGATATATTCACAGAAAGTACT 360
DB 301 GCATTGATCTTGAATATCATATTTTACTGACAGCCCTGATATATTCACAGAAAGTACT 360
QY 361 GAAACCGAAGACATCTCTGAAATGAGAAAAAACATTTCTTTACATTTCTGTGCCATACC 420
DB 361 GAAACCGAAGACATCTCTGAAATGAGAAAAAACATTTCTTTACATTTCTGTGCCATACC 420
QY 421 GAAAGTGAAGAAAGCCGAGTTAGGTTGATGCGAGATTTGTCAGTTGTTGCCAAGT 480
DB 421 GAAAGTGAAGAAAGCCGAGTTAGGTTGATGCGAGATTTGTCAGTTGTTGCCAAGT 480
QY 481 TTAATATCCCATTTTATCCCAATGTTTACAGTACGAATAAAGATATTTTGACACAGT 540
DB 481 TTAATATCCCATTTTATCCCAATGTTTACAGTACGAATAAAGATATTTTGACACAGT 540
QY 541 GGAAGATTTGACATTTGTTATGCCCCAAAAAGATAGATGATGCGGTTTGATGTTGCAACT 600
DB 541 GGAAGATTTGACATTTGTTATGCCCCAAAAAGATAGATGATGCGGTTTGATGTTGCAACT 600
QY 601 GCAATTTATGCTCTGATTTGATATAGAAATTTTCAAGCAAGTTGATTAAGTACGTTT 660
DB 601 GCAATTTATGCTCTGATTTGATATAGAAATTTTCAAGCAAGTTGATTAAGTACGTTT 660
QY 661 CAGGTTCTAGAAAGTGATCTGAGAAAGTTCCCAAGAGTTGAAAAAATTTGATGCAAGT 720
DB 661 CAGGTTCTAGAAAGTGATCTGAGAAAGTTCCCAAGAGTTGAAAAAATTTGATGCAAGT 720
QY 721 AACTGGGAATTCAAACATGAAGAAAGTATCAATTAACATACGAAATCAAGTTATTAAGGCT 780
DB 721 AACTGGGAATTCAAACATGAAGAAAGTATCAATTAACATACGAAATCAAGTTATTAAGGCT 780
QY 781 GACGTTCAAGGATGCTGAGAAACACCCAAATTTGATACAGATCTCAATGGAAG 840
DB 781 GACGTTCAAGGATGCTGAGAAACACCCAAATTTGATACAGATCTCAATGGAAG 840
QY 841 GAAAAGCCAGAAAGAGCTCTGTTGTATGACAGCTTAATTAAGTCCCAATTTACAGTTT 900
DB 841 GAAAAGCCAGAAAGAGCTCTGTTGTATGACAGCTTAATTAAGTCCCAATTTACAGTTT 900
QY 901 ATGAAGAAATTTGAGGGAATGCTGGAATAATCACTAGACCCAGAGACTTAATTTAA 960
DB 901 ATGAAGAAATTTGAGGGAATGCTGGAATAATCACTAGACCCAGAGACTTAATTTAA 960
QY 961 GAGTTAGATCATTTCTGTGAGCCTTTGACTGTGCGATTAAGAAATCAGAAAGGGTTA 1020
DB 961 GAGTTAGATCATTTCTGTGAGCCTTTGACTGTGCGATTAAGAAATCAGAAAGGGTTA 1020

```

```

DB 961 GAGTTAGATCATTTCTGTGAGCCTTTGACTGTGCGATTAAGAAATCAGAAAGGGTTA 1020
QY 1021 CAGAGATTATACAAAAATCAGAGTTTCAATTTGAACCTGATGTCAAACCCAGTTGTG 1080
DB 1021 CAGAGATTATACAAAAATCAGAGTTTCAATTTGAACCTGATGTCAAACCCAGTTGTG 1080
QY 1081 GACCGTTTCAAGAGATTCCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
DB 1081 GACCGTTTCAAGAGATTCCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
QY 1141 GATGCAATAGCTGTATATGATGTTGAAAATCAAGTGGAAATTTTAAAGCAGAAACTCTT 1200
DB 1141 GATGCAATAGCTGTATATGATGTTGAAAATCAAGTGGAAATTTTAAAGCAGAAACTCTT 1200
QY 1201 GAAAATCCAGATTTATTTTCAATATGTTTACTGCGTTGATTTGGAAGCAACAGAAAGT 1260
DB 1201 GAAAATCCAGATTTATTTTCAATATGTTTACTGCGTTGATTTGGAAGCAACAGAAAGT 1260
QY 1261 GTACTTGAAAGAAAAACGAGAGACTATATAGGTTATTA 1299
DB 1261 GTACTTGAAAGAAAAACGAGAGACTATATAGGTTATTA 1299

```

```

RESULT 4
ABZ31736
ID ABZ31736 standard; DNA; 1299 BP.
XX
XX ABZ31736;
AC 30-JAN-2003 (first entry)
DT
DT 30-JAN-2003 (first entry)
XX
XX Candida albicans essential gene SEQ ID NO 6023.
DE
DE Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
KM signal transduction; DNA replication; cell division; growth;
KM proliferation; Candida albicans; fungicide; antifungal; gene; ss.
XX
XX Candida albicans.
OS
OS WO200253728-A2.
XX
XX 11-JUL-2002.
PD
PD 26-DEC-2001; 2001WO-US049486.
XX
XX 29-DEC-2000; 2000US-0259128P.
XX 20-FEB-2001; 2001US-00792024.
PR 22-AUG-2001; 2001US-0314050P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
PA Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL,
PI WPI, 2002-566694/60.
XX
XX P-PSDS; ABP73186.
DR
DR
XX
XX Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele of
PT a gene and placing other allele of the gene under conditional expression.
XX
XX Claim 37; SEQ ID NO 6023; 167bp + Sequence listing; English.
XX
XX The invention relates to constructing (M1) a strain of diploid fungal
XX cells in which both alleles of a gene are modified, comprising modifying
XX one allele by insertion or replacement by a cassette having an
XX expressible selectable marker and modifying other allele by
XX recombination, of a promoter replacement fragment with a heterologous
XX promoter, so that expression of the second allele is regulated by the
XX promoter. (M1) is useful for constructing a strain of diploid fungal
XX cells in which both alleles of a gene are modified. The diploid fungal
XX cells having both alleles modified are useful for identifying a gene that
XX is essential to the survival or growth of a fungus, a gene that
XX contributes to the virulence and/or pathogenicity of a fungus, a gene

```

CC that contributes to the resistance of a diploid fungus to an antifungal
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus
 CC and for identifying a therapeutic agent for treatment of a mammalian
 CC disease. (M1) is useful for identifying a compound which modulates the
 CC activity of a gene product, preferably enzymatic activity, carbon
 CC compound catabolism, biosynthetic, transporter, transcriptional,
 CC translational, signal transduction, DNA replication and cell division
 CC activity. The method is useful for identifying a compound having the
 CC ability to inhibit growth or proliferation of *C. albicans* cells and for
 CC treating infection by *C. albicans*. The present sequence is that of an
 CC essential *Candida albicans* gene used in the method of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office
 CC XX

Sequence 1299 BP; 439 A; 221 C; 271 G; 368 T; 0 U; 0 Other;

Query Match 99.0%; Score 1286.2; DB 6; Length 1299;

Best Local Similarity 99.4%; Pred. No. 0;

Matches 1291; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 AAGTCAAAAGCAATTAGTACCTGGAAAGCAATTTCTGCTGGATATTGGTTCTT 60
 DB 1 AAGTCAAAAGCAATTAGTACCTGGAAAGCAATTTCTGCTGGATATTGGTTCTT 60
 QY 61 GAGCCAAATTATGATGCTTATGTCAGCATTTGTCATCAAGATGATCAAGTTATACA 120
 DB 61 GAGCCAAATTATGATGCTTATGTCAGCATTTGTCATCAAGATGATCAAGTTATACA 120
 QY 121 CCAAAAGAACCAAGTTGAAAGATCTAGATCAAAATTTCTTACCCCAATTTGCAAC 180
 DB 121 CCAAAAGAACCAAGTTGAAAGATCTAGATCAAAATTTCTTACCCCAATTTGCAAC 180
 QY 121 CGAAGATGGGAATATCAATATCATCAATCAAGCAAGCAAGATGATCAAGCTC 240
 DB 181 GAGGATGGGAATATCAATATCATCAATCAAGCAAGCAAGATGATCAAGCTC 240
 QY 181 GAGGATGGGAATATCAATATCATCAATCAAGCAAGCAAGATGATCAAGCTC 240
 DB 181 GAGGATGGGAATATCAATATCATCAATCAAGCAAGCAAGATGATCAAGCTC 240
 QY 241 AATAATCCATTTTATAGAGCAATATATCATGTTTATGCTTATATTCACACCGAA 300
 DB 241 AATAATCCATTTTATAGAGCAATATATCATGTTTATGCTTATATTCACACCGAA 300
 QY 301 GCATTTGATCTTGAATCATTTATCTGAGACCTGATATCATTCACAAGAGATCT 360
 DB 301 GCATTTGATCTTGAATCATTTATTTACTGAGACCTGATATCATTCACAAGAGATCT 360
 QY 361 GAAACCAAGACATCCCGAATGAGAAACATTTCTTACATCTCTGTCATTAAC 420
 DB 361 GAAACCAAGACATCCCGAATGAGAAACATTTCTTACATCTCTGTCATTAAC 420
 QY 421 GAAATGAAAGAACCGGATTTAGGTTTATGCGAGATTTGTCATGTTTGCACAAGT 480
 DB 421 GAAATGAAAGAACCGGATTTAGGTTTATGCGAGATTTGTCATGTTTGCACAAGT 480
 QY 481 TATATATCCATTTTATCCCAATGTTATCAAGTAAAGATATTTTGCACAAGT 540
 DB 481 TATATATCCATTTTATCCCAATGTTATCAAGTAAAGATATTTTGCACAAGT 540
 QY 541 GCAAGATTTGCAATGTTTATGCGCAAAAAAGATGAGATCTGCGTTTGAATGCAACT 600
 DB 541 GCAAGATTTGCAATGTTTATGCGCAAAAAAGATGAGATCTGCGTTTGAATGCAACT 600
 QY 601 GCAATTTATGCTGATGTTATATAGAAATTTCAAGCTTTGATTAATGACGTTT 660
 DB 601 GCAATTTATGCTGATGTTATATAGAAATTTCAAGCTTTGATTAATGACGTTT 660
 QY 661 CAGGTTCTAGAAAGTATCTGGAAGTTCCCAAGAGTGAAGAAATTTGATGAAGT 720
 DB 661 CAGGTTCTAGAAAGTATCTGGAAGTTCCCAAGAGTGAAGAAATTTGATGAAGT 720
 QY 721 AACTGGGAATTCAGAAATGATGATACCTTACCAAGGATCAAGTTATTAATGGGT 780
 DB 721 AACTGGGAATTCAGAAATGATGATACCTTACCAAGGATCAAGTTATTAATGGGT 780

QY 781 GAGTCAGAGGTGGCTCAGAAAACCCCAATTGTTATCAGATCTCAATGGAAG 840
 DB 781 GAGTCAGAGGTGGCTCAGAAAACCCCAATTGTTATCAGATCTCAATGGAAG 840
 QY 841 GAAAGCCAGAAAGAACTCTGTTGTTGATGACAGCTTAATGAGCCAAATTTACGTT 900
 DB 841 GAAAGCCAGAAAGAACTCTGTTGTTGATGACAGCTTAATGAGCCAAATTTACGTT 900
 QY 901 ATGAAGGAATTTAGGGAATGCGTAAATATGACACTCAGACCCAGACTTATATTA 960
 DB 901 ATGAAGGAATTTAGGGAATGCGTAAATATGACACTCAGACCCAGACTTATATTA 960
 QY 961 GAGTTAGATCATTTCTGTTGAGCTTTGACTGTTGCGATTAAGAACATCAAGAAAGGTTA 1020
 DB 961 GAGTTAGATCATTTCTGTTGAGCTTTGACTGTTGCGATTAAGAACATCAAGAAAGGTTA 1020
 QY 1021 CAGCATTTACCAAAATATCAGATTCATTTGAAACCTGATGTCAAACCCAGTTGTTG 1080
 DB 1021 CAGCATTTACCAAAATATCAGATTCATTTGAAACCTGATGTCAAACCCAGTTGTTG 1080
 QY 1081 GACCGTTGCAAGATTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1140
 DB 1081 GACCGTTGCAAGATTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1140
 QY 1141 GATGCAATGCTGATTTAGTTGTTGAAATCAAGTGGGAAATTTTAAAGCAAACTCTT 1200
 DB 1141 GATGCAATGCTGATTTAGTTGTTGAAATCAAGTGGGAAATTTTAAAGCAAACTCTT 1200
 QY 1201 GAAATCCGATTTATTTTATATATTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1260
 DB 1201 GAAATCCGATTTATTTTATATATTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1260
 QY 1261 GTACTTGAAGAAACCAAGAACTATATAGTTTATTA 1299
 DB 1261 GTACTTGAAGAAACCAAGAACTATATAGTTTATTA 1299

RESULT 5

AAD02788/c
 ID AAD02788 standard; DNA; 577 BP.

XX AAD02788;
 AC 31-MAY-2001 (first entry)
 DT 31-MAY-2001 (first entry)
 DE S. cerevisiae ERG8 gene homologous DNA #2 from *Candida albicans*.
 XX Phosphomevalonate kinase; PKM; ERG8; anti-fungal agent; diagnosis;
 KW infection; ds.
 OS *Candida albicans*.
 XX
 XX WO200114533-A2.
 EN 01-MAR-2001.
 PD 15-AUG-2000; 2000WO-GB003100.
 PF 21-AUG-1999; 99GB-00019766.
 PR (ASTR) ASTRAZENECA AB.
 PA (ASTR) ASTRAZENECA UK LTD.
 XX
 PI Rosamond JDC, Schnell NF;
 XX
 XX WPI; 2001-218441/22.
 DR New polypeptides and polynucleotides (ERG8) from *Candida albicans*, useful
 PT in assays for identifying inhibitors of phosphomevalonate kinase activity
 PT and as reagents for diagnosing *C. albicans* infection.
 XX
 XX Claim 6; Page 25; 29pp; English.

CC The patent discloses phosphomevalonate kinase (PMK; ERG8) protein and
 CC their corresponding DNA from Candida albicans. The ERG8 protein is
 CC useful in assays for identifying compounds that inhibit phospho-
 CC mevalonate kinase (PMK) activity. These inhibitors are useful as anti-
 CC fungal agents. The ERG8 DNA and protein are also useful as reagents for
 CC diagnosing C. albicans infection. The present sequence is a Candida
 CC albicans DNA which is homologous to the Saccharomyces cerevisiae ERG8
 CC gene

XX Sequence 577 BP, 167 A; 109 C; 108 G; 190 T; 0 U; 3 Other;

Query Match 40.1%; Score 521; DB 4; Length 577;
 Best Local Similarity 97.1%; Pred. No. 2,6e-129;

Matches 561; Conservative 0; Mismatches 13; Indels 4; Gaps 3;

QY 177 AAACGGAGAAATGGGAATATCATATCATCAATCAATACAGAACCCAGAGAATTGACGC 236
 DB 577 AACGGGGAATGGGAATATCATATCATCAATCAATACAGAAA-CCCAAGAGATGACGC 519
 QY 237 ACCGATTAAT-CCATTTTAAAGGCAACTATATTCATCG--TTTAACTTAATTCACCC 293
 DB 518 ACCGATTAATCCCATTTTAAAGGCAACTATATTCATCGGTTTAACTTAATTCACCC 459
 QY 294 GACCGAAGCATTTGATCTGAATCATCATCTTACTGACACCCCTGGATATCATTCACAGA 353
 DB 458 GACCGAAGCATTTGATCTGAATCATCATCTTACTGACACCCCTGGATATCATTCACAGA 399
 QY 354 AGATACTGAAACCAAGACATCTCGAATGAGAAAAAATTTCTTACATTCCTGATGC 413
 DB 398 AGATACTGAAACCAAGACATCTCGAATGAGAAAAAATTTCTTACATTCCTGATGC 339
 QY 414 CATTACCGAAGTGAAGAAACCGGATTAAGTTGATCGGACGATTAAGTTGATCGG 473
 DB 338 CATTACCGAAGTGAAGAAACCGGATTAAGTTGATCGGACGATTAAGTTGATCGG 279
 QY 474 CACAAGTTATATCCCATTTTATCCCAATGTTATCAGTACGAAATTAAGATATTGCA 533
 DB 278 CACAAGTTATATCCCATTTTATCCCAATGTTATCAGTACGAAATTAAGATATTGCA 219
 QY 534 CAAAGTTGACAGATTTGACATTTGTTATGCCAAAAAAAGATGATCGGTTGATGAT 593
 DB 218 CAAAGTTGACAGATTTGACATTTGTTATGCCAAAAAAAGATGATCGGTTGATGAT 159
 QY 594 TCGAATGCAATTTATGATCTGATGATATGAAAGATTTGACGACGTTGATTAATGA 653
 DB 158 TCGAATGCAATTTATGATCTGATGATATGAAAGATTTGACGACGTTGATTAATGA 99
 QY 654 CGTGTTCAGGTTCTAGAAAAGTATCTGAGAAAGTTCCCAAGAGTTGAAAAAATTGAT 713
 DB 98 CGTGTTCAGGTTCTAGAAAAGTATCTGAGAAAGTTCCCAAGAGTTGAAAAAATTGAT 39
 QY 714 TGAAGTAACTGGGAATTCAAACATGAAAAGATGTCAT 751
 DB 38 TGAAGTAACTGGGAATTCAAACATGAAAAGATGTCAT 1

RESULT 6
 AAD02787
 ID AAD02787 standard; DNA; 547 BP.

XX AAD02787;
 DT 31-MAY-2001 (first entry)

DE S. cerevisiae ERG8 gene homologous DNA #1 from Candida albicans.

KW Phosphomevalonate kinase; PMK; ERG8; anti-fungal agent; diagnosis;
 KM infection; ds.

XX Candida albicans.

XX OS
 XX PN
 XX MO000114533-A2.

PD 01-MAR-2001.

XX 15-AUG-2000; 2000WO-GB003100.

XX 21-AUG-1999; 99GB-00019766.

PA (ASTR) ASTRAZENECA AB.

PA (ASTR) ASTRAZENECA UK LTD.

XX Rosamond UDC, Schnell NF;

XX WPI; 2001-218441/22.

PT New polypeptides and polynucleotides (ERG8) from Candida albicans, useful
 PT in assays for identifying inhibitors of phosphomevalonate kinase activity
 PT and as reagents for diagnosing C. albicans infection.

XX Claim 6; Page 24; 29pp; English.

CC The patent discloses phosphomevalonate kinase (PMK; ERG8) protein and
 CC their corresponding DNA from Candida albicans. The ERG8 protein is
 CC useful in assays for identifying compounds that inhibit phospho-
 CC mevalonate kinase (PMK) activity. These inhibitors are useful as anti-
 CC fungal agents. The ERG8 DNA and protein are also useful as reagents for
 CC diagnosing C. albicans infection. The present sequence is a Candida
 CC albicans DNA which is homologous to the Saccharomyces cerevisiae ERG8
 CC gene

SQ Sequence 547 BP, 184 A; 81 C; 123 G; 159 T; 0 U; 0 Other;

Query Match 36.2%; Score 470.4; DB 4; Length 547;
 Best Local Similarity 99.8%; Pred. No. 9.5e-116;

Matches 471; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 828 CCAATGAAAAAGAAAAAGCCAGAAAGCTCTGTTGTGATGACCAAGTTAATGTC 887
 DB 1 CCAATGAAAAAGAAAAAGCCAGAAAGCTCTGTTGTGATGACCAAGTTAATGTC 60
 QY 888 CAATTTACAGTTATGAAGGAATGAGGGAATGCGTGAATAATACGATTCAGACCCAGA 947
 DB 61 CAATTTACAGTTATGAGGAATGAGGGAATGCGTGAATAATACGATTCAGACCCAGA 120
 QY 948 GACTTATATTAAGAGTTAGATCATCTGTTGAGCTTTGACGTTGATTAAGAACT 1007
 DB 121 GACTTATATTAAGAGTTAGATCATCTGTTGAGCTTTGACGTTGATTAAGAACT 180
 QY 1008 CAGAAAAAGGTTCAAGCATTAACAAAAATGAGGTTCCAAATGAACTGATGTCA 1067
 DB 181 CAGAAAAAGGTTCAAGCATTAACAAAAATGAGGTTCCAAATGAACTGATGTCA 240
 QY 1068 AACCCAGTTGTTGACCGTTGCAAGAGATTCCTGTTGTTGTTGTTGTTGTTCCAG 1127
 DB 241 AACCCAGTTGTTGACCGTTGCAAGAGATTCCTGTTGTTGTTGTTGTTGTTCCAG 300
 QY 1128 TGCTGTGATACGATGCAATAGCTGATTAAGTTGAAAAATCAAGTGGAAAAATTTAA 1187
 DB 301 TGCTGTGATACGATGCAATAGCTGATTAAGTTGAAAAATCAAGTGGAAAAATTTAA 360
 QY 1188 GCAGAAAACCTTTGAAAATCCGATTAATTTTCAATAAGTTTCTGGTTGATTTGGAAGA 1247
 DB 361 GCAGAAAACCTTTGAAAATCCGATTAATTTTCAATAAGTTTCTGGTTGATTTGGAAGA 420
 QY 1248 GCAGAACAGAGGTGATCTTGAAGAAAAACAGAAAGCTATATAGTTTATTA 1299
 DB 421 GCAGAACAGAGGTGATCTTGAAGAAAAACAGAAAGCTATATAGTTTATTA 472

RESULT 7
 AAL40805
 ID AAL40805 standard; DNA; 1356 BP.

XX AAL40805;

PT providing transformed cells having increased isoprenoid production.
 XX Disclosure; Page 116-117; 193pp; English.
 XX The invention relates to the use of specific genes of the mevalonate and
 CC isoprenoid biosynthetic pathways and inactive gene sites (pseudogenes).
 CC Genes of the invention are used to enhance biosynthesis of isopentenyl
 CC diphosphate (IP), dimethylallyl diphosphate (DMAPP) and isoprenoid
 CC pathway derived products in the plastids of transgenic plants and
 CC microalgae, for producing herbicide or antibiotic resistant transgenic
 CC plants and microalgae, for providing transformed cells with increased
 CC isoprenoid production compared to non-transformed cells, and for
 CC providing a cell with an inserted polynucleotide sequence encoding one or
 CC more products of interest. The present sequence is yeast
 CC phosphomevalonate kinase (PMK; ERG8) EC 2.7.4.2. encoding orf

XX Sequence 1356 BP; 427 A; 236 C; 304 G; 389 T; 0 U; 0 Other;

Query Match 10.2%; Score 132; DB 6; Length 1356;
 Best Local Similarity 50.5%; Pred. No. 8.1e-25;

Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCAAAAGCAATTTAGTCACCTGGAAAAGCAATTTCTTGCTGGGATATTTGGTCTTGA 62
 Db 9 GTTGAGAGCCTTGAGTCCCGGAGAAAGCTTACTGCTGGGATATTTAGTTTAA 68
 QY 63 GCCATTTATGATGCTTATGTACAGCATTTGTCATCAGATGCATGCAGTTTAAACCC 122
 Db 69 TACAAATATGAAACATTTGTAGTCGATATTCGCAAGAAATCATGCTGTAGCCATCC 128
 QY 123 AAAAGAAC-----CAGTTGAAAGATCTAATCAAAATTTCTTCAACCCCATTTGC 176
 Db 129 TTACGGTTCATTCAGAGGCTCTGATTAAGTTGAAGTCGTGAAAGTAAACAATTVA 188
 QY 177 AAACGAGAAATGGGAATATCATATCATCAATACAG--AGAAAGCCAGAGAAAGTTCA 233
 Db 189 AGATGGGAGTGGCTGTACATATAGCTCTAAAGTGGCTTCACTTCTGTTTCATAG 248
 QY 234 GTACCGATTAATCCATTTTGAAGCACTATATTCATGCTTTAGCTTATATTTCAACC 293
 Db 249 CGGATCTAAGAACCTTCTCATTAAGAAAGTTACGTAACGTATTTAGCTTAAACC 308
 QY 294 GACCGAGCATTTGATCTTGAAA-----TCATCATTTTACAGACCCGAGATATCA 344
 Db 309 TTAACATGACGACTACTGCAATAGAAACTGTTCTGTATGATATTTCTCGATGATGC 368
 QY 345 TTCACAGAGAACTGCAACCAAGACATCTCGAATGAGAAAACATTTCTTTACA 404
 Db 369 CTACCATTTCTCAGAGAGATAGGTTACCGAACATCGTGCAACAGAAATTTGAGTTTCA 428
 QY 405 TTCTGTCGCAATTAACGAGTGAAGAAAGCCGGAATTAAGTTCATGCGAGATTAAGTTC 464
 Db 429 TTCGACAGAAATTTGAAGAAAGTTCCCAAAAGAGGCTGCGCTCTCGAGAGTTTACAC 488
 QY 465 AATTGTGCGCAACAATTTATTAATCCATTTATCC-----CAATGTTATCAGTAC 515
 Db 489 AATTTTAATCAAGTTTGGCTCTCTTTTGTATCGGACCTGGAATAATATGAGCA 548
 QY 516 GAATAAATATTTTGGCAACAGTTGCAACAGTTGCAATTTGCTGATTTATAGAAATTTCA 575
 Db 549 ATATAGAGAAAGTTATCTAATTTAGCAAGATGCTCATTTGTCAAGCTCAGGTTAAAT 608
 QY 576 AGGATCTGGGTTGATGTTGCAACGCAATTTATGCTGATTTATATAGAAATTTCA 635
 Db 609 TGGAGCGGGTTTGTAGTAGCGGCGGAGCATATGATCTATCAGATTAAGAAATTTCC 668
 QY 636 GCCAGCTTGAATATGACGTTTCAGGTTCTGAAAAGTATCTTGAAGATTTCC 695
 Db 669 ACCGCAATTAATCTTAATTTTGCAGAAAT-----GAAAGTCTACTTACGCGAGTAA 722
 QY 696 AGAGTTGAAAAAATTTGATTAAGTAACTGGGAATTTCAAACTGAAGAAATGATACATTACC 755
 Db 723 ACTGGCGCATTTGGTTGATGAAGAAAGCTGAATATTTACATTAAGAAATGATACATTACC 782

QY 756 ATACGGAATCAAGTTATTAATGAGTACGTCAGAGGTGCTCAGAAACCAAAATTGCT 815
 Db 783 TTCGGATTAATCTTATGAGATGGGCAATTAAGAAATGTTCAAGAAAGTAAATCTGCT 842
 QY 816 ATCAGAGATATCTCCATGCAAGAAAGCCAGAAAGCTCTGTTGATGACCA 875
 Db 843 CCAAGAGGTAAAGAAATTTGATGATTCGCAATATGCAGAAAGCTTGAATAATATACGA 902
 QY 876 GCTTAATAGTCCATTTACAGTTTATGAGAAATTT 911
 Db 903 ACTGCATCATGCAATTTCTAATTTATGATGAGACT 938

RESULT 9

ABK96801
 ID ABK96801 standard; DNA; 1356 BP.

AC ABK96801;

XX 24-SEP-2002 (first entry)

DE 5. cerevisiae prenyl diphosphate synthase gene #20.

KW Prenyl alcohol; prenyl diphosphate synthase; geranylgeraniol;
 KW hydroxymethylglutaryl-CoA reductase; farnesylgeraniol; gene; ss;
 KW isopentenyl diphosphate delta-isomerase; mevalonate kinase;
 KW mevalonate CoA acetyltransferase; isoprenoid-terpenoid compound.

OS Saccharomyces cerevisiae.

PN WO200253746-A1.

PN 11-JUL-2002.

PF 20-DEC-2001; 2001WC-JP011214.

PR 28-DEC-2000; 2000JP-00403067.

PA (TOYT) TOYOTA JIDOSHA KK.

PI Ohto C, Obata S, Muramatsu M, Nishi K, Totusuka K;

DR WPI; 2002-537944/57.

PT Production of prenyl alcohols by culturing a transformant transferred
 PT with e.g. prenyl diphosphate synthase gene, for use in industrial
 PT synthesis of e.g. physiologically-active isoprenoid-terpenoid compounds.

PS Example 17; Page 269-270; 335pp; Japanese.

XX The invention relates to a process for producing a prenyl alcohol
 CC comprising: (A) construction of a recombinant by transferring an
 CC expression recombinant DNA or a DNA for genome integration into a host
 CC which contains prenyl diphosphate synthase gene or its variant; and (B)
 CC collecting product from the culture medium. Also described is: (1) a
 CC method for producing prenyl alcohol in which the expression recombinant
 CC DNA contains: (a) a hydroxymethylglutaryl-CoA reductase gene or its
 CC variant; or (b) an isopentenyl diphosphate delta-isomerase gene; (2) a
 CC process for producing geranylgeraniol, comprising: (a) constructing a
 CC recombinant by transferring an expression recombinant DNA or a DNA for
 CC genome integration into a host which contains hydroxymethylglutaryl-CoA
 CC reductase gene or its variant; and (b) isolating the product; (3) a
 CC process for producing farnesylgeraniol in which the expression
 CC recombinant DNA also contains an isopentenyl diphosphate delta-isomerase
 CC gene, mevalonate CoA acetyltransferase gene, hydroxymethylglutaryl-CoA
 CC synthase gene, mevalonate kinase gene, or mevalonate diphosphate
 CC decarboxylase gene. The methods are used for the production of prenyl
 CC alcohols, particularly for use in industrial synthesis of isoprenoid-
 CC terpenoid compounds e.g. physiologically-active prenyl alcohols including
 CC geometric isomers. ABK96780-ABK96897 represent prenyl diphosphate
 CC synthase genes and related PCR primers used in the methods of the
 CC invention

XX Sequence 1356 BP; 427 A; 236 C; 304 G; 389 T; 0 U; 0 Other;

Query Match 10.2%; Score 132; DB 6; Length 1356;
Best Local Similarity 50.5%; Pred. No. 8.1e-25;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

```

QY 3 GTCAAGAGCATTTAGTGCACCTGGAAGAACATTTCTGCTGGTGTGATATTTGGTCTTGA 62
DB 9 GTTAGAGGCTTCAAGTGGCCCAAGGAAAGCTTACAGTGTGGATATTAGTTTAA 68
QY 63 GCCAATTATGATGCTTATGACAGCATTTGCATCAAGATGCATGACAGTTATTAACCC 122
DB 69 TACAAATATGAAAGCATTTGTGTCGATTAATCGGCAAGAAATGCAATGCTGTAGCCATCC 128
QY 123 AAAAGAAC-----CAGTTGAAGATCTAGAAATCAAAATTTCTTCAACCCCAATTTCG 176
DB 129 TTACGGTTATTGCAAGGGTCTGATTAAGTTGAAGTGGCTGTGAAGAAATTAACAATTAA 188
QY 177 AAACGAGATGGGAATATCATATATCATCAATAACAG---AGAAAGCCAGAGAAAGTTCA 233
DB 189 AGATGGGAGTGGCTGTACATATAGTCTTAAGTGGCTTCAATTCCTGTTCCATAG 248
QY 224 GTCAGGCATTAATCCATTTTAAAGGCAATATATTCATGCTTTAGCTTAATTAATCAACC 293
DB 249 CGGATCTAAGAACCTTTTCATGAAAGATATCGCTAAGTATTAAGCTTAACTTTAAAC 308
QY 294 GACCGAGATTTGATCTTGAA-----TCATCATTTACTGACAGCCCTGATATCA 344
DB 309 TAAACATGACGACTACTGCAATGAAGAACTGTGCTGATTAATTTCTGATGATGC 368
QY 345 TTCACAAAGATGATCTGAAGAACGATCTCGAATGAGAAAGAAATTTCTTTACCA 404
DB 369 CTACATTTCTGAGAGATAGGCTTACCAACATGCTGCAACAGAAATTAAGTTTCA 428
QY 405 TTCTGTGTCATTAACGGAAGTGAAGAAAGCCGATTAAGTTGATCGGACAGATTAAGTGC 464
DB 429 TTCGACAGAAATGAAGAGTTCCCAAAACAGGGCTGGCTCTCGGACGTTTAACTAC 488
QY 445 AGTTGTGTCACAAAGTTTATATCCATTTTATCC-----CAATGTTATCACTAC 515
DB 489 AGTTTAACTACAGCTTGGCTCTCTTTTGTATCGAATCGGAAGAAATTAATGTAGACA 548
QY 516 GAATTAAGATATTTTGCACAAAGTTCACAGATTCGATGTAATGCCCAAAAAAGAT 575
DB 549 AATATGAGAGATTAATTCATTAATTTAGCACAAGTTGCTCATTTGCAAGCTCAAGGTAAT 608
QY 576 AGGATCTGGGTTGATGTTGCAACTGCAATTTATGCTGATTTGATATAGAAATTTCA 635
DB 609 TGGAGCGGGTTGATGTAAGCGGCGGACATATGATCTATCAGATATGAAGATTTCC 668
QY 636 GCCAGCTTGAATTAATGACGTGTTTCAGGTTCTGAAGATGATCTGAGAAAGTTCCCGAC 695
DB 669 ACCCCCATTAATCTCTAATTTGCCAGATATTT-----GGAAGGCTACTTACGGCAGTA 722
QY 696 AGAGTTGAAGAAATATGATTAAGTAAGTGAAGTTCGGAATTTCAAAATGAAGTATTAAC 755
DB 723 ACTGGGCAATTTGTTGATGAAGAAAGTGAATTAATTAAGTAAAGTAAACATTTAC 782
QY 756 ATACGGAATCAAGTATTAATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 815
DB 783 TTCCGGAATTAATTAATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 842
QY 816 ATCAGAGTATCTCAATGGAAGAAAGAAAGCCAGAAAGAGCTCTGTTGTTATGACCA 875
DB 843 CCAGAAATATGAAGAAATTTGATGATGATTCGATATGCCAGAAAGCTGGAAGAAATATATACAG 902
QY 876 GCTTATAGTGCATTTTACAGTTTATGAAGAAAT 911
DB 903 ACTGATCATGCAAAATTTTGAATTAATGATGACT 938

```

RESULT 10

```

ADP23287
ID ADP23287 standard; DNA; 1356 BP.
XX
XX AC ADP23287;
XX
XX DT 12-FEB-2004 (first entry)
XX
XX DE MevB operon phosphomevalonate kinase (PMK) gene.
XX
XX KM isopentenyl pyrophosphate; mevalonate pathway;
XX KM isopentenyl pyrophosphate synthesis; isoprenoid; pharmaceutical;
XX KM nutraceutical; flavouring agent; agricultural pest control agent;
XX KM acetoacetyl-CoA thiolase; atob; Mevalonate bottom operon; MevB operon;
XX KM gene; ds; phosphomevalonate kinase; PMK.
XX
XX OS Bacteria.
XX
XX PN US2003148479-A1.
XX
XX PD 07-AUG-2003.
XX
XX PF 06-DEC-2001; 2001US-00006909.
XX
XX PR 06-DEC-2001; 2001US-00006909.
XX
XX PA (KEAS/) KEASLING J.
XX PA (MART/) MARTIN V.
XX PA (PITE/) PITERA D.
XX PA (KIMS/) KIM S.
XX PA (WITH/) WITHERS S T.
XX PA (YOSH/) YOSHIKUNI Y.
XX PA (NEWM/) NEWMAN J.
XX PA (KHLE/) KHELEBNIKOV A V.
XX
XX PI Keasling J, Martin V, Pitera D, Kim S, Withers ST, Yoshikuni Y;
XX PI Newman J, Khelebnikov AV;
XX
XX DR WPI; 2004-020454/02.
XX
XX PT Synthesizing isopentenyl pyrophosphate in a host microorganism, useful
XX PT for pharmaceutical purposes, comprises introducing into the microorganism
XX PT heterologous nucleic acid sequences coding for an enzyme in the
XX PT mevalonate pathway.
XX
XX PS Claim 11; SEQ ID NO 5; 40pp; English.
XX
XX CC The invention describes a method of synthesizing isopentenyl
XX CC pyrophosphate (1) in a host microorganism. The method comprises
XX CC introducing into the host microorganism a plurality of heterologous
XX CC nucleic acid sequences each coding for a different enzyme in the
XX CC mevalonate pathway for producing isopentenyl pyrophosphate. The method is
XX CC useful in synthesizing isopentenyl pyrophosphate and the derived
XX CC isoprenoids. The isoprenoids may be used in pharmaceuticals,
XX CC nutraceuticals, flavouring agents or in agricultural pest control agents.
XX CC This sequence represents the phosphomevalonate kinase (PMK) gene from the
XX CC Mevalonate bottom (MevB) operon. Note: The specification states that MevB
XX CC and MevT operon gene are isolated from Saccharomyces cerevisiae and
XX CC Escherichia coli but does not state which species this polynucleotide was
XX CC isolated from.
XX
XX SQ Sequence 1356 BP; 426 A; 236 C; 305 G; 389 T; 0 U; 0 Other;
XX
XX Query Match 10.2%; Score 132; DB 12; Length 1356;
XX Best Local Similarity 50.5%; Pred. No. 8.1e-25;
XX Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;
XX
QY 3 GTCAAGAGCATTTAGTGCACCTGGAAGAACATTTCTGCTGGTGTGATATTTGGTCTTGA 62
DB 9 GTTAGAGGCTTCAAGTGGCCCAAGGAAAGCTTACAGTGTGGATATTAGTTTAA 68
QY 63 GCCAATTATGATGCTTATGACAGCATTTGCATCAAGATGCATGACAGTTATTAACCC 122
DB 69 TACAAATATGAAAGCATTTGTGTCGATTAATCGGCAAGAAATGCAATGCTGTAGCCATCC 128

```



```

QY 123 AAAAGAAC-----CAGTTGAAAGATCTAGAAATTTCTTACCCCAATTTGC 176
DB 129 TTACGGTTCAATGCAAGGGTCTGATTAAGTTGAAGTCCGTGTAAAGTAACATTTAA 188
QY 177 AAACGGAATGGGAATATCAATATCATTAATACAG--AGAACCCAGAGAGTTCA 233
DB 189 AGATGGGAGTGGCTGTACCATATTAAGTCTAAAGGGCTTCATTCGTGTTGATAG 248
QY 234 GTACGCATTAATCTATTTTAAAGGCACTATTAATCATGCTTTTGTAGTTTCAAC 293
DB 249 CGGATCTAAGAACCTTTCTTCAAAAAGTTATCGCTAACGATTAATTAAGCTTAAC 308
QY 294 GACCGAAGCATTTGATCTTGAAA-----TCATCATTTACTGAGCCCTGATATCA 344
DB 309 TTAACATGACGACTACGCAATGAAAATGTTGTTGATTAATTTTCTGATGATGC 368
QY 345 TTGCAAGAAGATACGTAACCAAGCATCTCGAATGAGAAAACATTTCTTTACA 404
DB 369 CTACCATTTCTCAGAGAGATAGCGTTACGAAACATCTGTGCAACAGAGATTGAGTTCA 428
QY 405 TTCTCGTCATTAACCGAAGTGGAAAAGACCGGATTAAGTTATCGGAGATTAGTGC 464
DB 429 TTGCGACAGATTTGAAGAGTTCCCAAAACAGGGCTGGCTCTCGGACAGTTTATGTCAC 488
QY 465 AGTTGTCGCAAGATTATATCCATTTTATCC-----CAATGTTATCAGTAC 515
DB 489 AGTTTAACTACAGCTTTGGCCCTCTTTTGTATGAGCATGGAATAATATGTAACA 548
QY 516 GAATTAAGATTTTGGCACAACGTTGCAAGATGCACTTTTATGCCCCAAAAGAT 575
DB 549 ATATAGAGATTAATTCATTAATTTAGCACAAGTTGCTCATTTGTCAAGCTCAGGTAAT 608
QY 576 AGGATCTGGTTGATGTTGCAACTGCAATTTATGCTGATTTGATTAAGAATTTCA 635
DB 609 TGAAGCGGGTTTGAATGTAAGCGCGGACACATATGATATACATATAGAAATTTCC 668
QY 636 GCCAGTTTGAATGAACGTGTTTCAAGTTCTAAGAAAGTATCTGAGAGTTCCCAAC 695
DB 669 ACCCGATTAATCTTAATTTGCCAGATATTT-----GGAAATGCTACTTACGGAGTAA 722
QY 696 AGAGTTGAAAAAATGATGAAAGTAAGTAAGTCAAAACATGAAAGTATACATTAC 755
DB 723 ACTGGCGCATTTGGTTGATGAAGAAAGCTGGAATATTCGATTTAAAGTAACCATTTAC 782
QY 756 ATACGGAATCAAGTTATTAATGGTGAACGTCAAGGTTGGCTCAGAAACCCCAATTTGAT 815
DB 783 TTCGGGATTAATTTATGATGGGCGATTTTAAGATGTTTCAAGAAACGTAATTAAGTGT 842
QY 816 ATCAGAGTACTCCATGGAAGAAAGAAAGCCAGAGAGAGCTCTGTTGTATGACCA 875
DB 843 CCAGAAAGTAAATAATGGTATGATTCGATATGCGCAGAAAGCTGAAATAATATATACGA 902
QY 876 GCTTAATAGTGCATTTTACAGTTTATGAAGGAAT 911
DB 903 ACTGATCATGCAATTTCTAAGTTATGATGACT 938

```

RESULT 11

AD182315
ID AD182315 standard; DNA; 1356 BP.

AD182315;

22-APR-2004 (first entry)

S cerevisiae phosphomevalonate kinase (PMK) gene.

amorph-4,11-diene synthesis; mevalonate pathway;

isopentenyl pyrophosphate; amorph-4,11-diene synthase gene;

pharmaceutical; nutraceutical; flavouring agent; agricultural pest control agent; epi-cedrol; phosphomevalonate kinase; PMK; gene; de.

```

XX OS Saccharomyces cerevisiae.
XX US2004005678-A1.
XX 08-JAN-2004.
XX 09-APR-2003; 2003US-00411066.
XX 06-DEC-2001; 2001US-00006909.
XX (KEAS/) KEASLING J.
XX (MART/) MARTIN V.
XX (PIRE/) PITERA D.
XX (WIT/) WITHERS S T.
XX (NEWM/) NEWMAN J.
XX Keasling J, Martin V, Pitera D, Withers ST, Newman J;
XX WPI; 2004-120864/12.
XX
XX This invention relates to a novel method of synthesizing amorph-4,11-
XX diene in a host microorganism which comprises introducing into the host
XX microorganism heterologous nucleic acid sequences, each coding for a
XX different enzyme in the mevalonate pathway for producing isopentenyl
XX pyrophosphate, and introducing into the host microorganism a DNA fragment
XX coding for an optimised synthetic amorph-4,11-diene synthase gene. The
XX nucleic acid sequences, DNA fragments, vectors and host cells are useful
XX for synthesizing isopentenyl pyrophosphate, amorph-4,11-diene and epi-
XX cedrol used as pharmaceuticals, nutraceuticals, flavouring agents, or
XX agricultural pest control agents. The present sequence is that of a gene,
XX encoding an enzyme, which was used during the exemplification of the
XX invention to assemble an operon for the synthesis of amorph-4,11-diene
XX in a host microorganism in the method of the invention.
XX
XX Sequence 1356 BP; 426 A; 236 C; 305 G; 389 T; 0 U; 0 Other;
XX
XX Query Match 10.2%; Score 132; DB 12; Length 1356;
XX Best Local Similarity 50.5%; Pred. No. 8.1e-25;
XX Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

```


Db	359	CTACATTCTCAGGAGGATGAGGTTACCGAACAATCGTGGCAACAGAAAGATTGAGTTTCA	428
Qy	405	TTCTGTCGCATTACCGAAGTGGAAAAAGACCGGATTAGGTTCAATCGGCAGAAATTAGTGC	464
Db	429	TTCCGACAGAAATTGAAGAAAGTTCCCAAAAACAGGCGTGGGCTCTCCGGCAGGTTAGTCA	488
Qy	465	AGTTGTTGCCACAAAGTTTATTTATCCCATTTTATCCC-----CAATGTTATCACTAC	515
Db	489	AGTTTAACTACAGCTTTGGCCCTCTTTTGTGATGGAAGCTGGAAAATATATCTAGCAA	548
Qy	516	GAATTAAGATATTTTGGCAACAAGTGGCAGACAGATGTGCAATGTATTATGCCAAAAAAGAT	575
Db	549	ATATGAGAGAGTTATTCATTAATTTAGCACAAAGTTGCTCATTTGCAAGCTCAGGGTAAAT	608
Qy	576	AGGATCTGGGTTTGATGTTGCAACTGCAAATTTATGCTGATTTGTAATTAGAAAGTTTCA	635
Db	609	TGGAAAGCGGGTTGATGTAAGTCGGCGGCAGACATATGATCTATTCAGATATATGAAGATTCCC	668
Qy	636	GCCAGCTTTGATTAATGACGTCGTTCAGGTTCTAGAAAGTATCTTGAGAAAGTTCCCCAC	695
Db	669	ACCCGCATTAATCTCTAAATTTCCAGATATTT-----GGAAAGCTACTTACCGCAGTAA	722
Qy	696	AGAGTTGAAAAAATTTGATGGAAGTACCTGGCAATTCAAACATGAAAGATGTACATTACC	755
Db	723	ACTGGCGGCTTTGGTTGATGAAAGAAAGACTGGAAATATTACGATTTAAAGTAAACCATTTACC	782
Qy	756	ATACGGAATCAAGATTATTATGAGTGACGTCGAAGGCTGCTCGAAGAACCCAAATTTGT	815
Db	783	TTCCGGAATTAACCTTTATGATGCGGCATATTTAAGATGCTTCAGAAAACAGTAAACTGGT	842
Qy	816	ATCAGCAGTACTCCAAATGGAAGAAAGAAAGCCAGAAAGAAAGCTGTGTGTATGACCA	875
Db	843	CCAGAAAGTTAAAAAATTTGGTATGATTCGATATGCGAGAAAGCTTGAAAAATATATACGA	902
Qy	876	GCTTAATAGTGCCAATTTCAGTTTATGAAAGAAAT 911	
Db	903	ACTGCATCATGCAAAATTTCAAGATTATGATGAGTACT 938	

RESULT 12
ADL14834
ID ADL14834 standard; DNA; 1356 BP.
XX
XX
AC ADL14834;
XX
DT 17-JUN-2004 (first entry)
XX
DE Yeast phosphomevalonate kinase gene ERG8.
XX
KM Yeast; ds; gene; Isopentenyl diphosphate biosynthetic pathway; IPP;
XX KM prenyl alcohol; squalene synthase.
OS Saccharomyces cerevisiae.
XX
PN US2004063182-A1.
XX
PD 01-APR-2004.
XX
PF 18-JUN-2003; 2003US-00450941.
XX
XX 28-DEC-2000; 2000JP-00401701.
PR 28-DEC-2000; 2000JP-00406067.
PR 18-SEP-2001; 2001JP-00282978.
PR 20-DEC-2001; 2001WO-JP011215.
XX
XX (OHTO/) OHTO C.
PA (OBAT/) OBATA S.
XX
XX Ohto C, Obata S;
XX
XX WPI; 2004-294401/27.
XX

Query Match	10.2%	Score 132	DB 12	Length 1356
Best Local Similarity	50.5%	Pred. No. 8.1e-25		
Matches 473	Conservative 0	Mismatches 430	Indels 33	Gaps 5
QY	3	GTCAAAGCATTGATGCACTGGAAGAAAGATTTCTTGCTGCTGATATTTGGTCTTGA	62	
DB	9	GTGAGAGCCTTCAGTGCCCAAGGAAAGCTTACTAGCTGTGATATTTAGTTTGA	68	
QY	63	GCCAAATTATGATCTTATGTGAAGCATTTGTCTATCAGATGCAAGTATTAACACC	122	
DB	69	TACAAATATATACCATTTGTAGTCGATTATCGGCAAGATGATGCTGATAGCCATCC	128	
QY	123	AAAAGAAC-----CAGTTGAAAGATCTAGAAATCAAATTTCTTACCCCAATTTGC	176	
DB	129	TTAGCGTTTCATTCGAAGGGCTGTGATTAAGTTGAAGTGGCTGTGAAGATAAACAATTTAA	188	
QY	177	AAAGGGAATGGGAATATATCATATATATCAAAATACAG--AGAAAGCCAGGAAGTTCA	233	
DB	189	AGATGGGAGTGGCTGTACATATATAGTCTTAAAGTGGCTTCAATCTGTGTTGCATAGG	248	
QY	234	GTCAAGCATTAATCCATTTTATAGAGCAATATATATTCATCGTTTATGCTTATATTAAC	293	
DB	249	CGATCTTAAGAACCTTTCATTTGAAAAAGTATGCTAAAGTATTTAGCTATTTAAAC	308	
QY	294	GACCGAAGCATTTGATCTTGAAA-----TATATTTTACTACAGACCCTTGATATCA	344	
DB	309	TAAATGAGACGATCTATGCAATTAAGAACTTGTTGGTTATTAATATTTCTCTGATGATGC	368	

QY 345 TTCAAGAAAGATCTGAAACCAAGACATCTCGAATGAGAGAAAAACATTTCTTACCA 404
 Db 369 CTACCAATTTCTCAGAGAGATAGCGTTACCGAACAATGTCGACAGAAAGATTAGTTTCA 428
 QY 405 TTCTCGGCAATTAACGGAAGTGGAAAAAGCCGATTAGTTTCATGCGGAGATTAGTGTG 464
 Db 429 TTGCGACAGAAATTTGAAGAAAGTTCGCCAAAACAGGGCTGCTCTCGGAGGTTTGTGAC 488
 QY 465 AGTTGTGGCCACAAAGTTTATATATCCATTTATGCC-----CAATGTTATCAGATAC 515
 Db 489 AGTTTAACTACAGCTTTGCGCTCTTTTGTATCGAACCTCGAAAAATTAATGACAA 548
 QY 516 GAATTAAGATTTTTCGACAGAGTTGACAGATTGACATTTGTTATGCCAAAAAAAGAT 575
 Db 549 AATTAAGAAATTTATTCATTAATTTAGACAAAGTTGCTATTTGCAAGCTCAGGGTAAAT 608
 QY 576 AGGATCTGGGTTTGAATGTTGCAACTGCAATTTATGCTGTATGTTATTAAGAAATTTCA 635
 Db 609 TGGAAAGCGGGTTTGAATGTTAGCGGGCGACGATATGATCTATCAGATATAGAAATTTCC 668
 QY 636 GCGAGCTTTGATTAATGACGTTGTTACAGTTCTAGAAAAGTATCTGAGAAAGTTCCCGAC 695
 Db 669 ACCGCAATTAATCTCTAATTTGCGCAGATATT-----GAAAGTCTACTTAACGCGAGTAA 722
 QY 696 AGAGTTGAAAAAATGATTTGAAGAATCTGGGAATTCGAACATGAAGAATGATCAATTACC 755
 Db 723 ACTGGCGCATTTGTTGATGTAAGAAAGATGGAATTAATTAAGAAATTAATTAATTAATTA 782
 QY 756 ATAGGAAATCAATTTATTAATGAGTGAAGTGAAGGTTGCTCAGAAAACCCCAAAATTTGAT 815
 Db 783 TTGCGGATTAATCTTATGAGATGCGCGATTTAAGAAATGTTTCAAGAAACAGTAAATCTGGT 842
 QY 816 ATCAGAGTACTCCAAATGGAAGAAAGAAAGCCAGAAAGAAAGTCTGTTGTGTATGACCA 875
 Db 843 CCAGAAAGTAAAAAATTTGATGATGATTCGATATGCGCAAGAAAGCTGAAAAATATATACAGA 902
 QY 876 GCTTAATAGTCCCAATTTTACAGTTTATGAGAAGAAAT 911
 Db 903 ACTGATCATGCAAAATTTCTAGATTTTATGATGATGACT 938

RESULT 13

ADP23291
 ID ADP23291 standard; DNA; 4482 BP.

AC ADP23291;

DT 12-FEB-2004 (first entry)

DE Mevalonate pathway operon MevB (Mevalonate bottom).

XX isopentenyl pyrophosphate; mevalonate pathway;
 KM isopentenyl pyrophosphate synthesis; isoprenoid; pharmaceutical;
 KM nutritional; flavouring agent; agricultural pest control agent;
 KM acetate:CoA thiolase; atob; Mevalonate bottom operon; MevB operon;
 KM gene; db; mevalonate kinase; MK; phosphomevalonate kinase; FMK;
 XX mevalonate pyrophosphate decarboxylase; MPD.

OS Bacteria.

PN US2003148479-A1.

PD 07-AUG-2003.

PF 06-DEC-2001; 2001US-00006909.

PR 06-DEC-2001; 2001US-00006909.

XX (KEAS// KEASLING J.
 PA (MART// MARTIN V.
 PA (PITE// PITERA D.
 PA (KIMS// KIM S.
 PA (WITH// WITHERS S T.

PA (YOSH// YOSHIKUNI Y.
 PA (NEM// NEMMAN J.
 PA (KHE// KHEBNIKOV A V.
 XX
 PI Keasling J, Martin V, Pitera D, Kim S, WITHERS ST, Yoshikuni Y;
 PI Newman J, Khebnikov AV;
 XX
 DR WPI; 2004-020454/02.

PT Synthesizing isopentenyl pyrophosphate in a host microorganism, useful
 PT for pharmaceutical purposes, comprises introducing into the microorganism
 PT heterologous nucleic acid sequences coding for an enzyme in the
 PT mevalonate pathway.

XX Claim 55; SEQ ID NO 9; 40bp; English.

XX The invention describes a method of synthesizing isopentenyl
 CC pyrophosphate (i) in a host microorganism. The method comprises
 CC introducing into the host microorganism a plurality of heterologous
 CC nucleic acid sequences each coding for a different enzyme in the
 CC mevalonate pathway for producing isopentenyl pyrophosphate, and the
 CC useful in synthesizing isopentenyl pyrophosphate and the derived
 CC isoprenoids. The isoprenoids may be used in pharmaceuticals and
 CC nutraceuticals, flavouring agents or in agricultural pest control agents.
 CC This sequence represents the mevalonate pathway operon Mevalonate bottom
 CC (MevB) which contains genes encoding mevalonate kinase (MK),
 CC phosphomevalonate kinase (PMK) and mevalonate pyrophosphate decarboxylase
 CC (MPD). Note: The specification states that MevB and MevT operon gene are
 CC isolated from *Saccharomyces cerevisiae* and *Escherichia coli* but does not
 CC state which species this polynucleotide was isolated from.

Seq Sequence 4482 BP; 1337 A; 928 C; 972 G; 1245 T; 0 U; 0 Other;

Query Match 10.2%; Score 132; DB 12; Length 4482;
 Best Local Similarity 50.5%; Pred. No. 1.3e-24;

Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCAAAAGCATTTAGTGAAGCACTGGAAGAAAGCATTTCTGCTGTGATTAATTTGTTCTTGA 62
 Db 1606 GTTGAAGCCTTATGATGAGTCCCGGAGAAAGCTTACTAGCTGTGATTAATTTGTTTGA 1665
 QY 63 GCCAATTTATGATGCTTATGTAAGCATGTCATCAGAAATGATGATGATTAATTAACACC 122
 Db 1666 TACAAAATATGAAGCATTTGTAGTGAATTAATGCGCAAGATGATGATGATGATGATG 1725
 QY 123 AAAAGAAC-----CAGTTGAAGATCTAGATCAAAAATTTCTTCAACCCCAATTTGC 176
 Db 1726 TTACGCTTATGTAAGGAGTGTGATTAATTTGAAGTGTGTAAGAAAGTAAACATTTAA 1785
 QY 177 AAAAGGAGATGGAATATCATATATCAATATCAAAATACAG---AGAAAGCCAGAGAAATTC 233
 Db 1786 AGATGGGAGGATGGCTGTACATATATAGCTTAAGAGTGGCTTCAATTCCTGTTGGAATGG 1845
 QY 234 GTACAGCAATTAATTCATTTTATGAGGCACTATATTCATGCTTTTATGCTTATTAACCC 293
 Db 1846 CGAATCTAAGAACCTTTTCATTTGAAGAAAGTATGCTTAAGCTATTTAGCTAATTTAAAC 1905
 QY 294 GACCGAAGCATTTGATCTTGA---TCATGATTTATCTCAACCTGATATCA 344
 Db 1906 TAAATGAGACGATCTGCAATATAGAACTGTTGTTGATTAATTTCTCTGATGATGTC 1965
 QY 345 TTCAAGAAAGATCTGAAACCAAGACATCTCGAATGAGAGAAAAACATTTCTTTTACCA 404
 Db 1966 CTACCATTTCTGAGAGATGCTTTACCGAACATCGTGGCAAGAAAGATTTGATTTTCA 2025
 QY 405 TTCTCGTCCATTTCCGAGTGGAAAAAGACCGGATTAATGTTTATCGGACAGATTAATGTTGTC 464
 Db 2026 TTGCGACAGAAATTTAAAGAAAGTTCGCAAAACAGGCTGCTCTCGGAGGTTTATGATCAC 2085
 QY 465 AGTTGTGGCCACAAAGTTTATATCCATTTATGCC-----CAATGTTATCAGTAC 515
 Db 2086 AGTTTAACTACAGCTTTGCGCTCTTTTGTATGAGACCTGAAAAATTAATGATGACAA 2145

QY 516 GAATAAGATATTTGACACAGCTTGACAGATTGCACTTTATGCCCAAAAAAGAT 575
 DB 2146 ATATAGAGAGTTTATTCATTAATTTAGCAACAGTTGCTCAAGTCAAGGTTAAAT 2205
 QY 576 AGGATCTGGGTTGATGTTGCACTGCAATTATGCTGATGTTATAGAAAGATTGA 635
 DB 2206 TGGAAAGCGGTTGATGTAAGCGCGACGACATATGATCTATCAGATTAAGAAGATTCCC 2265
 QY 636 GCCAGCTTTGATTAATGACGTGTTTCAGGTTCTAGAAAGTATCCGAGAAGTCCCCAC 695
 DB 2266 ACCCGCATTAATCTCTAATTTCCAGATATTT-----GGAAGTCTACTTACGGCAGTAA 2319
 QY 696 AGAGTTGAAAAAATGATTTGAAAGTAACTGGAATTTCAACATGAAGAATGTACATTACC 755
 DB 2320 ACTGGCGCATTTGGTTGATGAAAGAGCTGGAATTTAGATTTAAAGTAACATTACC 2379
 QY 756 ATACGAATCAAGTTATTTATGAGGTGACCTCAAGGTTGCTCAAAAACCCCAATTGCT 815
 DB 2380 TTCGGGATTTAATCTTATGATGCGGATTTTAAGATGCTTCAGAAACAGTAAACTGGT 2439
 QY 816 ATCAGAGTACTCCAAATGGAAGAAAGAAAGCAAGAAAGCTGTTGTGATGACCA 875
 DB 2440 CCAGAAAGTTAAAAATTTGTTATGATTCGATATGCCAGAAAGCTGAAAAATATATACAA 2499
 QY 876 GCTTAATAGTCCAAATTTACAGTTTATGAAGAATT 911
 DB 2500 ACTGCATCATGCAATTTCTAGATTTATGATGATGACT 2535

RESULT 14

AD182319
 ID AD182319 standard; DNA; 4482 BP.

AC AD182319;

DT 22-APR-2004 (first entry)

DE Mevalonate bottom (MeVB) operon DNA sequence.

KM amorph-4,11-diene synthase; mevalonate pathway;
 KM isopentenyl pyrophosphate; amorph-4,11-diene synthase gene;
 KM pharmaceutic; nutraceutical; flavouring agent;
 KM agricultural pest control agent; epi-cedrol; Mevalonate bottom operon;
 KM MeVB; ds.

OS Saccharomycetes cerevisiae.
 OS Synthetic.

PN US2004005678-A1.

PD 08-JAN-2004.

PF 09-APR-2003; 2003US-00411066.

PR 06-DEC-2001; 2001US-00006909.

PA (KEAS/) KEASLING J.
 PA (MART/) MARTIN V.
 PA (PITE/) PITERA D.
 PA (WITH/) WITHERS S T.
 PA (NEWM/) NEWMAN J.

PI Keasling J, Martin V, Pitera D, Withers ST, Newman J;

DR WPI; 2004-120864/12.

PT Synthesizing amorph-4,11-diene in a host cell, useful as
 PT pharmaceuticals, comprises introducing nucleic acid sequences, each
 PT coding for a different enzyme in the mevalonate pathway for producing
 PT isopentenyl pyrophosphate.

PS Example 1; SEQ ID NO 9; 75bp; English.

CC This invention relates to a novel method of synthesizing amorph-4,11-
 CC diene in a host microorganism which comprises introducing into the host
 CC microorganism heterologous nucleic acid sequences, each coding for a
 CC different enzyme in the mevalonate pathway for producing isopentenyl
 CC pyrophosphate, and introducing into the host microorganism a DNA fragment
 CC coding for an optimised synthetic amorph-4,11-diene synthase gene. The
 CC nucleic acid sequences, DNA fragments, vectors and host cells are useful
 CC for synthesizing isopentenyl pyrophosphate, amorph-4,11-diene and epi-
 CC cedrol used as pharmaceuticals, nutraceuticals, flavouring agents, or
 CC agricultural pest control agents. The present sequence is that of an
 CC operon which may be used in the method of the invention.

XX Sequence 4482 BP; 1337 A; 928 C; 972 G; 1245 T; 0 U; 0 Other;

SQ Query Match 10.2%; Score 132; DB 12; Length 4482;

Best Local Similarity 50.5%; Pred. No. 1.3e-24;

Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCAAAAGCATTTAGTGCACCTGAAAGCATTTCTGTGTGTGATATTTGTTTGA 62
 DB 1606 GTTGAGAGCTTCAGTGGCCCCAGGAAAGCGTTACTAGCTGTGATATTTAGTTTAA 1665
 QY 63 GCCAATTTATGATGCTTATGTAAGCATGTGTCATCAAGATGATGATTAACACC 122
 DB 1666 TACAAAATATGAAGCATTTGTAAGTGAATTAAGCAAGAAATGCTGTAGCCCATCC 1725
 QY 123 AAAAGAAC-----CAGTTGAAAGAAATCTGAATCAAAATTTCTTACCCCAATTTC 176
 DB 1726 TTACGTTTCATTCGAAGGCTCTGATTAAGTTTAAGTGCCTGTGAAAGTAACATTTAA 1785
 QY 177 AAACGAGAAATGGGAATATCAATATCAATCAATCAAG--AGAAAGCCAGAAAGTTCA 233
 DB 1786 AATGGGGAGTGGCTGTAACATATTAAGTCTTAAAGTGGCTTCATTCGTTTCAATAG 1845
 QY 234 GTACGCAATTAATCCATTTTATGAGCACTATATTCATGTTTATGCTTAATTTCAACC 293
 DB 1846 CGGATCTAAGAACCTTTTCATTAAGAAAGTTATGCTTAAGCTTAATTTAAMCC 1905
 QY 294 GACCGAAGCATTTGATCTTGAA-----TCATCATTTACTGACAGCCCTGATATCA 344
 DB 1906 TTAACATGACGACTACTGCAATTAAGAACTGTTGCTATGATTAATTTTCTGATGATGC 1965
 QY 345 TTCACAAGAAATGACTGAAACCAAGACATCTCGATGAGAGAAAGAAACATTTCTTTACCA 404
 DB 1966 CTACCATCTCGAGAGAGATAGGTTACGSAACATGCTGCAACAGAAAGTTGATTTTCA 2025
 QY 405 TTCTGTGCTATTACCGAAAGTGAAGACCGGATTTAGGTTATGCTGACGATTAAGTGTGC 464
 DB 2026 TTGCGACAGAAATGAAGAGTTCCCAAAAAGGGGCTGCGCTCTCGGCGAGTTTATGCTAC 2085
 QY 465 AGTTGTGCGACAAGTTTATTTATCCCATTTTATCCC-----CAATGTTATCAGTAC 515
 DB 2086 AGTTTAACTACAGCTTGGCCCTCTTTTGTATGCGACCTGGAAGAAATTAATGTAGCAA 2145
 QY 516 GAATAAGATATTTGACAAAGTTGACAGATTCACATTTGTTATGCCCAAAAAAGAT 575
 DB 2146 ATATAGAGAGTTTATTCATTAATTTAGCAACAGTTGCTCAAGTCAAGGTTAAAT 2205
 QY 576 AGGATCTGGGTTGATGTTGCACTGCAATTATGCTGATGTTATAGAAAGATTGA 635
 DB 2206 TGGAAAGCGGTTGATGTAAGCGCGACGACATATGATCTATCAGATTAAGAAGATTCCC 2265
 QY 636 GCCAGCTTTGATTAATGACGTGTTTCAGGTTCTAGAAAGTATCCGAGAAGTCCCCAC 695
 DB 2266 ACCCGCATTAATCTCTAATTTCCAGATATTT-----GGAAGTCTACTTACGGCAGTAA 2319
 QY 696 AGAGTTGAAAAAATGATTTGAAAGTAACTGGAATTTCAACATGAAGAATGTACATTACC 755
 DB 2320 ACTGGCGCATTTGGTTGATGAAAGAGCTGGAATTTAGATTTAAAGTAACATTACC 2379
 QY 756 ATACGAATCAAGTTATTTATGAGGTGACGTCAGAGGTGCTCAAGAAACCCCAATTGCT 815
 DB 2380 TTCGGGATTTAATCTTATGATGAGGCGATTTTAAGATGCTTCAAGAAACAGTAAACTGGT 2439

QY 816 ATCAGAGTCTCCATGAGAAAAGCAGAGAAAGCTCTGTGTGATGACCA 875
 DB 2440 CCAGAGGTAAAAATGGTATGATTCGATATGCGAAGAGCTTGAATAATATACGA 2499
 QY 876 GCTTAATAGTGCATTTTACAGTTATGAGAAATT 911
 DB 2500 ACTGATCATGCAAAATCTAGATTATGATGACT 2535

RESULT 15
 ADF23294
 ID ADF23294 standard; DNA; 5051 BP.
 AC ADF23294;
 XX
 DT 12-FEB-2004 (first entry)
 DE Recombinant MevB operon MBI.
 XX
 KW isopentenyl pyrophosphate; mevalonate pathway;
 KW isopentenyl pyrophosphate synthesis; isoprenoid; pharmaceutical;
 KW nutraceutical; flavouring agent; agricultural pest control agent;
 KW acetoacetyl-CoA thiolase; atob; Mevalonate bottom operon; MevB operon;
 KW gene; ds; isopentenyl pyrophosphate isomerase; idi;
 XX dimethylallyl pyrophosphate; DMAPP; IPP; MBI.
 OS Synthetic.
 OS Escherichia coli.
 OS Bacteria.
 XX
 FN US2003148479-A1.
 XX
 PD 07-AUG-2003.
 XX
 PF 06-DEC-2001; 2001US-00006909.
 XX
 PR 06-DEC-2001; 2001US-00006909.
 XX
 PA (KEAS/) KEASLING J.
 PA (MART/) MARTIN V.
 PA (PITER/) PITERA D.
 PA (KIMS/) KIM S.
 PA (WITTH/) WITHERS S T.
 PA (YOSH/) YOSHIKUNI Y.
 PA (NEWM/) NEWMAN J.
 PA (KHLE/) KHLEBNIKOV A V.
 XX
 PI Keasling J, Martin V, Pitera D, Kim S, Withers ST, Yoshikuni Y,
 PI Newman J, Khlebnikov AV;
 DR WPI; 2004-020454/02.
 XX
 PT Synthesizing isopentenyl pyrophosphate in a host microorganism, useful
 PT for pharmaceutical purposes, comprises introducing into the microorganism
 PT heterologous nucleic acid sequences coding for an enzyme in the
 PT mevalonate pathway.
 PS Example 1; SEQ ID NO 12; 40bp; English.
 XX
 CC The invention describes a method of synthesizing isopentenyl
 CC pyrophosphate (IP) in a host microorganism. The method comprises
 CC introducing into the host microorganism a plurality of heterologous
 CC nucleic acid sequences each coding for a different enzyme in the
 CC mevalonate pathway for producing isopentenyl pyrophosphate. The method is
 CC useful in synthesizing isopentenyl pyrophosphate and the derived
 CC isoprenoids. The isoprenoids may be used in pharmaceuticals,
 CC nutraceuticals, flavouring agents or in agricultural pest control agents.
 CC This sequence represents the MBI operon created from adding the E. coli
 CC isopentenyl pyrophosphate isomerase gene idi to the MevB operon to ensure
 CC sufficient production of dimethylallyl pyrophosphate (DMAPP) from
 CC isopentenyl pyrophosphate (IPP), required to create the backbone
 CC structure of all isoprenoids.

XX SQ Sequence 5051 BP; 1486 A; 1064 C; 1122 G; 1379 T; 0 U; 0 Other;
 Query Match 10.2%; Score 132; DB 12; Length 5051;
 Best Local Similarity 50.5%; Pred. No. 1.4e-24;
 Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;
 QY 3 GTCAAAAGCATTTAGTGAACCTGAGAAAGCATTTCTGCTGAGATATTTGGTTCTTGA 62
 DB 1606 GTTGAGAGCCTTCAGTCCCGGAGGAAAGCGTTACTGCTGTGATTTAGTTTGA 1665
 QY 63 GCCAATTTATGATGCTTTATGTCAGAGCATTTGTCTATCAGAAATGATGATTAACACC 122
 DB 1666 TACAAATATGAGCATTTATGATGCTGATTAATGCGCAAGATGATGCTGATGCCATTC 1725
 QY 123 AAAAGAAC-----CAGTTGAAAGATCTAGATCAAAATTTCTCACCCCAATTGC 176
 DB 1726 TTACGTTCAATTGCAAGGCTCTGATTAAGTTGAAGTCGCTGAAAGTAAACATTTAA 1785
 QY 177 AAACGGAATGGAATATCAATATCATCAATCAATACAG---AGAGCCAGAGAGTTCA 233
 DB 1786 AGATGGGAGTGGCTGTAACATATAGTCTTAAAGTGCTTCATCTCTGTTGATAGG 1845
 QY 234 GTCAAGCATTAATCATTTTATAGAGCAAGCATATTCATGTTTATGCTTATATCAACC 293
 DB 1846 CGGATCTAAGAACCTTTCAATTGAAAGATTAATGCTTACGATTTAGCTTAAACC 1905
 QY 294 GACGAAAGCATTTGATCTTGA---TCATCATTTTACTCAGACCCCTGATATCA 344
 DB 1906 TTAATGAGAGACTACGCAATGAAAGATTTGTTGTTATTAATTTCTCTGATATGC 1965
 QY 345 TTCACAGAAAGATCTGAAGCAAGCATCTCGAATGAGAAACATTTCTTATCA 404
 DB 1966 CTACCATTTCTCAGAGATATGCTTACCGAATCTGCGCAAGAGATGATGATTTCA 2025
 QY 405 TTCTGTCATTACCGAAGTGGAAAGACCGGATTTAGTTCATCGGAGATTTAGTTC 464
 DB 2026 TTGCAAGAAATGAAAGATTTCCCAAAACAGGGCTCTCGGAGGTTTATGTCAC 2085
 QY 465 AGTTGTTCCAGCAAGTTTATATCCATTTTATCC-----CAATGTTATCAGTAC 515
 DB 2086 AGTTTATCTACAGCTTTGGCTCTCTTTTGTATCGACCTGCAAAATTAATGTAGACA 2145
 QY 516 GAATTAAGATATTTTGCACAGCTTGCACAGATTGCAATGTTATGCCCAAAAGAT 575
 DB 2146 ATATAGAGAGTATTCATTAATTTAGCAAGATTTGCTCATTTGTCAAAGCTCAGGGTAAAT 2205
 QY 576 AGGATCTGGTTGATGTTGCACTGCAATTTATGCTGATTTGATATAGAAAGATTTCA 635
 DB 2206 TGAAGCGGTTGATGTAAGCGCGGAGCATATGATCTATCAGATATAGAAAGTTCC 2265
 QY 636 GCCAGCTTGATTAATGACGTGTTCAAGTCTTAAGAAATGATCTGGAAGTTCCAC 695
 DB 2266 ACCCGATTAATCTATATTTCCAGATATTT-----GAAAGTGTACTTTCGGCAGTAA 2319
 QY 696 AAGATTGAAAAAATGATTAAGTAAGTAAGTGAATTCAAACATGAAGATGTATACATTAAC 755
 DB 2320 ACTGGCGCATTTGTTGATGAAGAAAGACTGGAATATTCAGATTAAGTAAGCAATTAAC 2379
 QY 756 ATACGAATCAAGTATTAATGAGGAGCTCAAGGTTGCTCAGAAACCCCAATTTGT 815
 DB 2380 TTGGGATTAATCTTATGATGAGGAGATTAAGAAATGTTTCAAAAACAGTAAATCGGT 2439
 QY 816 ATCAGAGTCTCCATGAGAAAAGCAGAGAAAGCTCTGTGTGATGACCA 875
 DB 2440 CCAGAGGTAAAAATGGTATGATTCGATATGCGAAGAGCTTGAATAATATACGA 2499
 QY 876 GCTTAATAGTGCATTTTACAGTTATGAGAAATT 911
 DB 2500 ACTGATCATGCAAAATCTAGATTATGATGACT 2535

RESULT 16

AD182322
 ID AD182322 standard; DNA; 5051 BP.
 AC AD182322;
 XX
 XX 22-APR-2004 (first entry)
 DT
 DE Amorpha-4,11-diene synthetase-related MBI operon DNA sequence Segid12.
 XX
 XX isomorph-4,11-diene synthetase; mevalonate pathway;
 KM isopentenyl pyrophosphate; amorpha-4,11-diene synthase gene;
 KM pharmaceutical; nutraceutical; flavouring agent;
 KM agricultural pest control agent; epi-cedrol; MBI operon; ds.
 XX
 OS Saccharomyces cerevisiae.
 OS Escherichia coli.
 PN US2004005678-A1.
 XX
 XX 08-JAN-2004.
 PD
 XX 09-APR-2003; 2003US-00411066.
 PF
 XX 06-DEC-2001; 2001US-00006909.
 PR
 XX (KEAS/) KEASLING J.
 PA (MART/) MARTIN V.
 PA (PIRE/) PITERA D.
 PA (WITH/) WITHERS S. T.
 PA (NEWM/) NEWMAN J.
 XX
 PI Keasling J, Martin V, Pitera D, Withers ST, Newman J;
 DR WPI; 2004-120864/12.
 XX
 XX Synthesizing amorpha-4,11-diene in a host cell, useful as
 PT pharmaceuticals, comprises introducing nucleic acid sequences, each
 PT coding for a different enzyme in the mevalonate pathway for producing
 PT isopentenyl pyrophosphate.
 XX
 XX Example 1; SEQ ID NO 12; 75bp; English.
 PS
 CC This invention relates to a novel method of synthesizing amorpha-4,11-
 CC diene in a host microorganism which comprises introducing into the host
 CC microorganism heterologous nucleic acid sequences, each coding for a
 CC different enzyme in the mevalonate pathway for producing isopentenyl
 CC pyrophosphate, and introducing into the host microorganism a DNA fragment
 CC coding for an optimized synthetic amorpha-4,11-diene synthase gene. The
 CC nucleic acid sequences, DNA fragments, vectors and host cells are useful
 CC for synthesizing isopentenyl pyrophosphate, amorpha-4,11-diene and epi-
 CC cedrol used as pharmaceuticals, nutraceuticals, flavouring agents, or
 CC agricultural pest control agents. The present sequence is that of an
 CC operon which may be used in the method of the invention.
 CC
 SQ Sequence 5051 BP; 1486 A; 1064 C; 1122 G; 1379 T; 0 U; 0 Other;

Query Match 10.2%; Score 133; DB 12; Length 5051;
 Best Local Similarity 50.5%; Pred. No. 1.4e-24;
 Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCAAAAGCATTTAGTGCACCTGTAAGAAAGCATTTCTTGGTGTGATATTGCTTGA 62
 DB 1606 GTTGAGAGCTTCAGTGGCCCGAGGAAAGCGTTACTAGTGTGATATTGCTTTTAA 1665
 QY 63 GCCAATTATGATGCTTATGTGACAGCATTTGTCATCAAGATGATGACGATTATTAACCC 122
 DB 1666 TACAATAATGAAAGCATTTGTAGTCGATATCGGCAAGAAAGCATGCTGAGCCCATCC 1725
 QY 123 AAAAGAAC-----CAGTTGAAAGAAATCTGAATCAAAATTTCTTACCCCAATTGCG 176
 DB 1726 TTACGCTTATTCGAAAGGCTGTGATTAAGTTGAAGGCGGTGTAAGAAAGTAAACAAATTAA 1785
 QY 177 AAACGAGAAATGGAATATCATATCATCAATAACAG--AGAAAGCCAGAAAGTTCA 233

DB 1786 AGATGGGAGTGGCTGTATCCATATTAAGTCTTAAAGTGCTTCATTCCTGTTTCGATAG 1845
 QY 234 GTACAGCAATAATTCATTTTGTAGAGCAACTATATTCATCGTTTAGCTTATATCAACC 293
 DB 1846 CGGATCTAAGAACCTTTCATTAAGAAAGTTATCGCTTAACGTAATTAAGCTTATTAACC 1905
 QY 294 GACCGAAGCATTTGATCTTGAAA-----TCATCATTTTACTGAGACCTGATATCA 344
 DB 1906 TAAACATGACGACTACTGGAATATGAAACCTGTTCCTATATGATATTTCTCGATGATGC 1965
 QY 345 TTCACAAAGAAATGACTGAAACCAAGACATCTCGATGAGAAAGAAACATTTCTTACCA 404
 DB 1966 CTACCATTTCTCAGAGAGTAGCGTTACGAAACATGCTGCAACAGAAATAGATTTC 2025
 QY 405 TTCTGTCGCTATACCGAAGTGGAAAACCGGATTTAGTTGATGCGCAGATTAAGTGC 464
 DB 2026 TTTCGACGAATGGAAGAGTTCCCAAAACAGGCGCTCTCGCAGGTTTATGTCAC 2085
 QY 465 AGTTGTGCAACAGTTTATTTATCCCATTTTATCC-----CAATGTTATCAGTAC 515
 DB 2086 AGTTTATCTACAGCTTTGGCCTCTTTTGTATCGACCTGAAATATATGATAGACA 2145
 QY 516 GAATTAAGATATTTTGCACACAGCTTGACAGATGCAATTTGTATGCCAAAAAAGAT 575
 DB 2146 ATATAGAGAAGTTATTCATTAATTTAGCAAGTGTCTCATTTGCAAGCTCAGGTTAAAT 2205
 QY 576 AGGATCTGGGTTGATGATTTGCACTGCAATTTATGCTGATGATATTAAGAAATTTCA 635
 DB 2206 TGGAAAGCGGTTGATGTAAGCGCGGCGGCAATATGATCTATACAGATTAAGAAATTTCC 2265
 QY 636 GCCAGCTTTGATTAATGACAGTGTTCAGGTTCTAGAAAGTATCCTGAGAAATTCGCCAC 695
 DB 2266 ACCGCTTATCTCTAATTTTCCAGATATTT-----GGAAGTCTACTTACGCAATTA 2319
 QY 696 AGAGTTGAAAAAATTTGATTAAGTAACTGGAATTTCAAAACATGAAAGATGTACATTAC 755
 DB 2320 ACTGCGCATTTGTTGATTAAGAAAGACTGGAATTTACGATTAAGAAATTCATTAC 2379
 QY 756 ATACGGAATCAAGTTATTAATGCGTACGCTCAAGGCTGCTGAAACCCCAATTTGCT 815
 DB 2380 TTCCGATTTAATCTTATGATGAGGCGGATTAAGATGTTTCAAGAAACGTAATTTGCT 2439
 QY 816 ATCAAGAGTACTCCATGGAAGAAAGAAAGCCAGAAAGAAAGCTGTGTGTATGACCA 875
 DB 2440 CCAGAAAGTAAAGAAATTTGATGATTCGCAATGCGCAAGAAAGCTGAAATATATACGA 2499
 QY 876 GCTTAATAGTGCATTTTACATTTATGAAAGAAAT 911
 DB 2500 ACTGCATCATGCAAAATTTAGATTTATGATGATGACT 2535

RESULT 17

ADP23295
 ID ADP23295 standard; DNA; 5963 BP.

AC ADP23295;
 DT 12-FEB-2004 (first entry)
 XX
 XX
 DE Recombinant MevB operon MBIS.

isopentenyl pyrophosphate; mevalonate pathway;
 KM isopentenyl pyrophosphate synthetase; isoprenoid; pharmaceutical;
 KM nutraceutical; flavouring agent; agricultural pest control agent;
 KM acetoacetyl-CoA thiolase; atob; Mevalonate bottom operon; MevB operon;
 KM gene; ds; farnesyl pyrophosphate synthase; ispa; MBIS.

OS Synthetic.
 OS Escherichia coli.
 OS Bacteria.
 XX
 XX US2003148479-A1.

```

XX 07-AUG-2003.
PD
XX 06-DEC-2001; 2001US-00006909.
PF
XX 06-DEC-2001; 2001US-00006909.
PR
XX
PA (KEAS/) KEASLING J.
PA (MART/) MARTIN V.
PA (PITERA/) PITERA D.
PA (KIMS/) KIM S.
PA (WITHERS/) WITHERS S T.
PA (YOSH/) YOSHIKUNI Y.
PA (NEWMAN/) NEWMAN J.
PA (KHEBNIKOV/) KHEBNIKOV A V.
PI Keasling J, Martin V, Pitera D, Kim S, Withers ST, Yoshikuni Y;
PI Newman J, Khebnikov AV;
XX
XX WPI; 2004-020454/02.
DR
XX
XX Synthesizing isopentenyl pyrophosphate in a host microorganism, useful
PT for pharmaceutical purposes, comprises introducing into the microorganism
PT heterologous nucleic acid sequences coding for an enzyme in the
PT mevalonate pathway.
XX
XX Example 1; SEQ ID NO 13; 40bp; English.
PS
XX The invention describes a method of synthesizing isopentenyl
XX pyrophosphate (I) in a host microorganism. The method comprises
XX introducing into the host microorganism a plurality of heterologous
XX nucleic acid sequences each coding for a different enzyme in the
XX mevalonate pathway for producing isopentenyl pyrophosphate. The method is
XX useful in synthesizing isopentenyl pyrophosphate and the derived
XX isoprenoids. The isoprenoids may be used in pharmaceuticals,
XX nutraceuticals, flavouring agents or in agricultural pest control agents.
XX This sequence represents the MBIS operon comprising the E. coli farnesyl
XX pyrophosphate synthase gene ispA added to the MB1 operon which contains
XX the MevB operon plus the isopentenyl pyrophosphate isomerase gene to
XX direct the product of the mevalonate pathway operon.
XX
XX Sequence 5963 BP; 1700 A; 1305 C; 1374 G; 1584 T; 0 U; 0 Other;
SQ
Query Match 10 2%; Score 132; DB 12; Length 5963;
Best Local Similarity 50.5%; Pred. No. 1.5e-24;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;
QY 3 GTCAAAAGCATTTAGTGCACCTGAGAAAGCATTTCTTGCTGGTGGATATTTGGTTCTTGA 62
DB 1606 GTTGAGAGCTTCAAGTGGCCCGGAGAAAGCATTTAGTGGTGGATATTTGGTTTGA 1665
QY 63 GCCAATTTATGATGCTTATGTGACAGCATTTGCATCAGCAATGATGATGATTAATACAC 122
DB 1666 TACAAAATATGAAGCATTTGTAGTGGATTAATCGCAAGAAATGATGCTGATGCCATCC 1725
QY 123 AAAAGAAC-----CAGTTGAAGAATCTAGATCAAAATTTTTCACCCCAATTGGC 176
DB 1726 TTACGCTTCATTGCAAGGCTGATTAAGTTGAAGTGGTGGAGAAAGTAAACATTTTA 1785
QY 177 AAACGAGAAATGGGAATATCAATATCATCAATCAAG---AGAGCCAGAGAAAGTTCA 233
DB 1786 AGATGGGGAGTGGCTGATCAATATAGTCTTAAAGTGGCTTCACTCTGTTTGCATAGG 1845
QY 234 GTACGCGATTAATTCATTTTATGAGCAACTATATTCATCGTTTATGCTTATATTCAC 293
DB 1846 CGGATCTAAGAACCTTTTCATTTGAAAGATTAATCGTAACGATTTAGCTATTAAC 1905
QY 294 GACCGAAGCATTTGATCTTGAAA-----TATCATTTTATCAAGCCCTGATATCA 344
DB 1906 TAACTGAGACGACTCTGCAATAGAACTGTTGCTTAATGATATTTTCTCTGATGATGC 1965
QY 345 TTCAACAGAGATATCTGAACCAAGACATCTCGAATGAGAAAGAAACATTTCTTTACCA 404

```

```

DB 1966 CTACCATTTCTGAGAGATAGGCTTACCGAATCATGTCGACAGAGAAATGAGTTTCA 2025
QY 405 TTCTGTCGATTAACCGAAGTGGAAAAAGCCGATTTAGTTTCATGGCAGATTTATGTC 464
DB 2026 TTGCGACAGAAATGAGAGATTTCCCAAAAAGGGCTGGGCTCTCCGAGAGTTTATGTCAC 2085
QY 465 AGTTGTCGACAGATTTATATCCATTTTATCC-----CAATGTTATCAGTAC 515
DB 2086 AGTTTATCTACAGTTTGGCTCTCTTTTGTATGCACTCGAAGAAATATATGATGACA 2145
QY 516 GAATTAAGATATTTTGCACAAAGTTGCACAGATTCACATTTGTTATGCCCCAAAAAGAT 575
DB 2146 ATATAGAGAGATTTATCTAATTTATAGCACAAGTTGCTCATTTGTCAGGCTCAGGTTAAAT 2205
QY 576 AGGATCTGGGTTTGTATGTTGACATGCAATTTATGTCGTCTGTATATATGAAAGATTTCA 635
DB 2206 TGGAGCGGGTTTGTATGAGCGGCGAGCATATGATGATATATGATTAAGAAATTTCC 2265
QY 636 GCCAGCTTTGATTAATGACGCTGTTTCAAGTTCTAGAAAGTATCTGGAAGTTCCCCAC 695
DB 2266 ACCGCAATTAATCTCTAATTTTGCAGATTT-----GGAAGTCTACTTACGGCAGTAA 2319
QY 696 AGATTTGAAAAAATTTGATTGAAGTAACTGGGAATTCAGATGAAAGATGATACATTACC 755
DB 2320 ACTGCCGATTTGTGTATGATGAGAAAGACTGGAATATTTACATTAAGTAACCATTTAAC 2379
QY 756 ATACGAAATCAAGTTATTAATGAGTGAAGTCAAGAGGTGGCTCAGAAACACCAATTTGT 815
DB 2380 TTGCGGATTAATCTTATGATGAGCGCATTTATAGATTTGTCAGAAACAGTAAATCTGT 2439
QY 816 ATCAAGGATCTCCATGAGAAAGAAAGCCAGAAAGAAAGCTCTGTTGTATGACCA 875
DB 2440 CCAGAGGTAATAAATTTGATGATTCGATATGCGAATGCGAAAGCTTGAATATATATACGA 2499
QY 876 GCTTAATATGTCCAATTTTACAGTTTATGAAGAAAT 911
DB 2500 ACTGATCATGCAAAATTTCTAGATTTATGATGACT 2535

```

RESULT 18
AD182323
ID AD182323 standard; DNA; 5963 BP.
AC AD182323;
XX
XX 22-APR-2004 (first entry)
XX
XX Amorpha-4,11-diene synthase is-related MBIS operon DNA sequence SegID13.
XX
XX amorpha-4,11-diene synthase; mevalonate pathway;
XX isopentenyl pyrophosphate; amorpha-4,11-diene synthase gene;
XX pharmaceutical; nutraceutical; flavouring agent;
XX agricultural pest control agent; epi-cedrol; MBIS operon; ds.
XX
XX *Saccharomyces cerevisiae*.
OS *Escherichia coli*.
XX
XX US2004005678-A1.
XX
XX 08-JAN-2004.
XX
XX 09-APR-2003; 2003US-00411066.
XX
XX 06-DEC-2001; 2001US-00006909.
XX
XX (KEAS/) KEASLING J.
XX (MART/) MARTIN V.
XX (PITERA/) PITERA D.
XX (WITHERS/) WITHERS S T.
XX (NEWMAN/) NEWMAN J.
XX
XX Keasling J, Martin V, Pitera D, Withers ST, Newman J;

DR WPI, 2004-120864/12.
 XX Synthesizing amorpho-4,11-diene in a host cell, useful as
 PT pharmaceuticals, comprises introducing nucleic acid sequences, each
 PT coding for a different enzyme in the mevalonate pathway for producing
 PT isopentenyl pyrophosphate.
 PS Example 1; SEQ ID NO 13; 75bp; English.
 XX
 XX This invention relates to a novel method of synthesizing amorpho-4,11-
 CC diene in a host microorganism which comprises introducing into the host
 CC microorganism heterologous nucleic acid sequences, each coding for a
 CC different enzyme in the mevalonate pathway for producing isopentenyl
 CC pyrophosphate, and introducing into the host microorganism a DNA fragment
 CC coding for an optimised synthetic amorpho-4,11-diene synthase gene. The
 CC nucleic acid sequences, DNA fragments, vectors and host cells are useful
 CC for synthesizing isopentenyl pyrophosphate, amorpho-4,11-diene and epi-
 CC cedrol used as pharmaceuticals, nutraceuticals, flavouring agents, or
 CC agricultural pest control agents. The present sequence is that of an
 CC operon which may be used in the method of the invention.
 XX
 XX Sequence 5963 BP; 1700 A; 1305 C; 1374 G; 1584 T; 0 U; 0 Other;
 SQ

Query Match 10.2%; Score 132; DB 12; Length 5963;
 Best Local Similarity 50.5%; Pred. No. 1.5e-24;
 Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCAAAAGCATTTAGTGCACCTGGAAGCAATTTCTGCTGCGATATTGGTTCTTGA 62
 DB 1606 GTTGAGCGCTTCAGTGGCCCGAGGAAACGGTACTAGCTGTGGATTTGATTTTAA 1665
 QY 63 GCCAATTTATGATGCTTATGTGACAGCATTTGCATCAACGATGACAGTTATTAAC 122
 DB 1666 TACAAATATGAGCATTTGTAGTGGATTAACGGAAGATGCAATGCTGAGCCATCC 1725
 QY 123 AAAAGAAC-----CAGTTGAAGAAATCTAGAAATCAAAATTTCTTCACCCCTTGC 176
 DB 1726 TTACGCTTCATTCGAAAGGCTGTGATTAAGTTGAAGTCGTGAAAGTAAACAAATTTA 1785
 QY 177 AAACGAGAAATGGGAATATACATATCATCAATTAACAG--AGAACCCAGAGAAAGTTCA 233
 DB 1786 AGATGGGAGTGGCTGTACATATTAAGTCTTAAAGTGGCTTCATTCCTGTTTCATAG 1845
 QY 234 GTACAGCATTAATTCATTTTATGAGGCAATTAATTCATGCTTTTATAGTTATTTCAAC 293
 DB 1846 CGGATCTAAGAACCTTTTCATTTGAAAGATTAACGTAACGATTTTATGCTTTAAAC 1905
 QY 294 GACCGAAGCATTTGATCTTGA-----TCATCATTTTACTGACACCTGATATCA 344
 DB 1906 TAAACATGACGACTACTGCAATTAAGAACTTTCGTTATGATTAATTTCTGATGATGC 1965
 QY 345 TTCAAGAAAGTACTGAAACCAAGACATCTCGAATGAGAAAGAAATTTCTTTACA 404
 DB 1966 CTACCATTTCTGAGAGATAGGTTAACGACATCGTGGCAACAGAAAGTTAGTTTCA 2025
 QY 405 TTCTGTCGTCATTTACGAAAGTGAAGACCGGATTTAGTTGTCGCGAGATTAAGTGC 464
 DB 2026 TTTCGACAGAAATTTGAAGAGTTCCCAAAACAGGGCTGCGGCGTTTACTGC 2085
 QY 465 AGTTGTCGCAAGATTTATTTATCCATTTTATCC-----CAATGTTATCAGTAC 515
 DB 2086 AGTTTAACTACAGCTTTGGCTCTCTTTTGTATCGACCTGGAAGATTAATGTAGACA 2145
 QY 516 GAATTAAGATATTTTGCACACGTTGCAAGATTTGCACTTTGTATGCCCAAAAAAGT 575
 DB 2146 ATATGAGAGATTTATTCATTAATTTAGCAAGATTTGCTATTTCAAGCTCAGGGTAAAT 2205
 QY 576 AGGATCTGGGTTGATGTTGCAAGTCAATTTATGCTGATTTGATTAAGAAATTTCA 635
 DB 2206 TGGAAAGCGGTTGATGTCGCGGCGGACATATGATTTATCAATATTAAGAAATTTCC 2265
 QY 636 GCCACCTTTGATTAATGACGTGTTTCAGGTTCTAGAAAGTGAATCTGGAAGTTCCCCAC 695

DB 2266 ACCGCGATTAATCTCTTAATTTGCCAGATTT-----GGAGTCTTACTTACGGCAGTAA 2319
 QY 696 AGAGTTGAAAAAATTTGATTAAGATGAGTGGCAATTTCAACATGAAGATGTACATTACC 755
 DB 2320 ACTGGCGCATTTGTGTTGATTAAGAAAGCTGGAATTTTACGATTAAGAAATTTAC 2379
 QY 756 ATACGGAATCAAGTTATTAATGCTGAGCTCAAGGCTGCTCAGAAACCAACCAATTGGT 815
 DB 2380 TTCCGGATTTAACTTTATGATGATGGCGCATTTAAAGATGGTTTCAGAAACAGTAAAC 2439
 QY 816 ATCAGAGTATCTCCATGGAAGAAAGCAAGAAAGCTCTGTTGTTATGACCA 875
 DB 2440 CCAGAGATTAAGAAATTTGATTAATGCTGCAAGAAAGCTTGAATTAATACGA 2499
 QY 876 GCTTATAGTGGCAATTTTACAGTTTATGAAGGAATT 911
 DB 2500 ACTGCATCATGCAATTTCTAGATTTATGATGACT 2535

RESULT 19

AAD31026
 ID AAD31026 standard; DNA; 7681 BP.

XX AAD31026;
 XX 29-AUG-2003 (revised)
 DT 31-MAY-2002 (first entry)

DB Operon D DNA encoding mevalonate pathway.

XX Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;
 KW isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;
 KW transgenic plant; yeast; phosphomevalonate kinase; HMGR; ACT;
 KW mevalonate kinase; mevalonate diphosphate decarboxylase; HMG-CoA; MD;
 KW acetoacetyl thiolase; 3-hydroxy-3-methylglutaryl-coenzyme A synthase;
 KW MVK; PMK; HMG-CoA reductase; ds.

XX Saccharomyces cerevisiae.
 OS Arabidopsis thaliana.
 OS Streptomyces sp.
 OS Chimeric.

PN WO200210398-A2.

PD 07-FEB-2002.

PF 31-JUL-2001; 2001WO-US024037.

PR 31-JUL-2000; 2000US-0221703P.

PA (HANN/) HAHN F M.

PI (KUEH/) KUEHNIS A R.

PI Hahn FM, Kuehnle AR;

DR WPI; 2002-217122/27.

PT Use of specific genes of mevalonate and isoprenoid biosynthetic pathways,
 PT for providing a cell with herbicide or antibiotic resistance, and for
 PT providing transformed cells having increased isoprenoid production.

PS Claim 77; Page 141-145; 193bp; English.

XX The invention relates to the use of specific genes of the mevalonate and
 CC isoprenoid biosynthetic pathways and inactive gene sites (pseudogene).
 CC Genes of the invention are used to enhance biosyntheses of isopentenyl
 CC diphosphate (IPP), dimethylallyl diphosphate (DMAPP) and isoprenoid
 CC pathway derived products in the plastids of transgenic plants and
 CC microalgae, for producing herbicide or antibiotic resistant transgenic
 CC plants and microalgae, for providing transformed cells with increased
 CC isoprenoid production compared to non-transformed cells, and for
 CC providing a cell with an inserted polynucleotide sequence encoding one or
 CC more products of interest. The present sequence is operon D DNA encoding

CC the entire mevalonate pathway. This operon contains *S. cerevisiae* orf
CC encoding phosphomevalonate kinase (PMK), mevalonate kinase (MVK),
CC mevalonate diphosphate decarboxylase (MDP), acetosuccinyl thiolase (ACOT
CC and *A. thaliana* orfs encoding 3-hydroxy-3-methylglutaryl- coenzyme A (
CC-CoA) synthase (HMGs) and Streptomyces sp. CL130 orf encoding HMG-CoA
CC reductase (HMGrt). (Updated on 29-AUG-2003 to standardise OS field)
XX
SQ Sequence 7661 BP; 2127 A; 1691 C; 1837 G; 2026 T; 0 U; 0 Other;

Query Match	10.2%;	Score 132;	DB 6;	Length 7681;
Best Local Similarity	50.5%;	Pred. No. 1.6e-24;		
Matches 473;	Conservative	0;	Mismatches 430;	Indels 3

Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

OY	5	GTCAAAAGCATTTTGTGCACCTGGAAAAGCATTTCTTGTGTGTGATAATTGGTCTTGA	62
Db	43	GTTAGAGCCCTTCAGTGTCCCGAGGAAAGGCTTACGTGCTGGATATTTAGTTTAGA	102
OY	63	GCCAAATTATGATGCTATGTGACAGATTTGTGATACGAATGACATGATTAACACC	122
Db	103	TACAAATATGAAAGCATTTTGTAGTCGGATTTATGCGCAAGAAATGCAATGCTGAGCCCATCC	167
OY	123	AAAAGGAAAC-----CAGTTGAAAGATCTAGAAATCAAAATTTCTTCAACCCCAATTTGC	176
Db	163	TTACGGTTCATTTGCAAGGGCTGTGATTAAGTTTGAAGTGGCTGTGAAAGTAAACATTTTAA	222
OY	177	AAACGGAGATTTGGGAATATCATATTCATCAAAATTCAG---AGAGCCCAAGAAAGTTCA	233
Db	223	AGATGGGGAATGGCTGTACATATTAAGTCTTAAAGTGGCTTCATTTCTGTTTGGATAGG	282
OY	234	GTCACGATTAATTCATTTTATAGGCAACTAATTCATCGTTTAGCTTATATTAACCC	293
Db	283	CGGATCTTAAGAACCCCTTTCATTTGAAAAAGTATGCGTACAGATTTTACTACTTTAAACC	342
OY	294	GACCGAAGCATTTTGATCTTGAA-----TCATCATTTACTAGACCTCGATATCA	344
Db	343	TACATGAGACGACTCTGCATATGAAACTTGTTCGTATATGATATTTCTTGTATGATGC	402
OY	345	TTCAAGAAAGATATCTGAAACCAAGCATCTCGAATGAGAAAGAAACATTTCTTTACA	404
Db	403	CTACATTTCTCAGAGGATAGCGTTATCCGAATGCTGGCAACAGAAATTTGAATTTCA	462
OY	405	TTTCCGGGCATTTACCGAATGGAAAAAGCCGATTTAGTTTCATGCGCAGATTAAGTGC	464
Db	463	TTGCGACAGATTTGAAGAAAGTTCCTCAAAACAGGGCTGGGCTCTCTCGGAGGTTTAACTAC	522
OY	465	AGTTGTGCCAAGTTTATTTATCCATTTATGCC-----CAATGTTTACAGTAC	515
Db	523	AGTTTAACTACAGCTTTTGGCCCTCTTTTGTATCGAAGCTTGAAATTAATGTAGCA	582
OY	516	GAATTAAGATTTTGTGCACAGGTTGCACAGATTGCACATTTGTATGCCCAAAAAGAT	575
Db	583	ATATAGGAAGTATTCATTAATTTAGCACAAAGTGTCTATTTGCACACTCAGGGTAAAT	642
OY	576	AGGATCGGGTTTATGATGTGCAACTGCATTTATAGTGTGATATATAGAAAGTTTCA	635
Db	643	TGGAAGGGGTTTGAATGACGGCGGACGACATATGATCTATCAGATATTAAGAAATTC	702
OY	636	GCCAGCTTGATTAATGACGCTGTTCAAGGTTCTAGAAAGTATCTGAGAAGTTCCCA	695
Db	703	ACCGCATTAATCTCTAATTTGCCAGATATT-----GGAAGTCTACTTACGGCAGTAA	756
OY	696	AGAGTTGAAAAAATTTATGAAAGTATACGGGAATTCAAACATGAAGATGTACATTACC	755
Db	757	ACTGGGCAATTTGCTTGAATGAAGAACTGGAAATTTTCATTAAGTAACCATTTTACC	816
OY	756	ATACGGAATCAAGTTATTAATGGGTGACGTCAAGGGTGGCTCAGAAACAACCAATTTGT	815
Db	817	TTGCGGATTTAATCTTAATGATGGGCGATATTAAGAAATGTTTCAGAAACAGTAAACCTGT	876
OY	816	ATCAGCAATCTCAATGGAAAAAGGAAACCCAGAAAGAAAGCTCTGTGTGTATGACA	875
Db	877	CCGAAGGTAATAAATTTGGTATGATTTGCCATATGCGCAAGAAAGCTTGAAATTAATATACAG	936

Ox 876 GCTTAATAGTCCCAATTACAGTTTATGAAGAATT 911
||| ||| ||| ||| ||| ||| |||
Db 937 ACTGATCATGCAAATTCAGATTATGATGACT 972

RESULT 20
AAD31023
ID AAD31023 standard; DNA; 7693 BP.

AC AAD31023;

DT	29-AUG-2003	(revised)
DT	31-MAY-2002	(first entry)

Operon A DNA encoding mevalonate pathway.

KM Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP
KM isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;
KM transgenic plant; yeast; phosphomevalonate kinase; HMGAT; HMGCS; ACT;
KM mevalonate kinase; mevalonate diphosphate decarboxylase; HMG-CoA; MDD
KM acetoacetyl thiolase; 3-hydroxy-3-methylglutaryl-coenzyme A synthase;
KM MVK; PDK; HMG-CoA reductase; da.

OS *Saccharomyces cerevisiae*.
OS *Arabidopsis thaliana*.
OS Chimeric.

PN WO200210398-A2

PD 07-FEB-2002.

PF 31-JUL-2001; 2001WO-US024037.

PR 31-JUL-2000; 2000US-0221703P.

PA (HAHN/) HAHN F M.

XX

XX

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

PT for providing a cell with herbicide or antibiotic resistance, and for

[illegible]

33

CC The invention relates to the use of specific genes of the mevalonate and
CC isoprenoid biosynthetic pathways and inactive gene sites (pseudogenes).
CC Genes of the invention are used to enhance biosynthetic rates of isopentenyl
CC diphosphate (IPP), dimethylallyl diphosphate (DMAPP) and isoprenoid
CC pathway derived products in the plastids of transgenic plants and
CC microalgae, for producing herbicide or antibiotic resistant transgenic
CC plants and microalgae, for providing transformed cells with increased
CC isoprenoid production compared to non-transformed cells, and for
CC providing a cell with an inserted polynucleotide sequence encoding one or
CC more products of interest. The present sequence is operon A DNA encoding
CC the entire mevalonate pathway. This operon contains 5. cerevisiae orfs
CC encoding phosphomevalonate kinase (PMK), mevalonate kinase (MVK),
CC mevalonate diphosphate decarboxylase (MDD), acetosuccinyl thiolase (AACT)
CC and A. thaliana orfs encoding 3-hydroxy-3-methylglutaryl- coenzyme A (HMG-
CC CoA) synthase (HMGGS) and HMG-CoA reductase (HMGGR). (Updated on 29-AUG-
CC 2003 to standardise OS field)

SQ Sequence 7693 BP; 2212 A; 1526 C; 1794 G; 2161 T; 0 U; 0 Other;

Query Match	10.28;	Score 132;	DB 6;	Length 7693;
-------------	--------	------------	-------	--------------

Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCAAAGCATTTAGTGCACCTGAAAAGCATTTCTGCTGTCGATATTTGGTTCCTGA 62

Db	37196	GTGTAGAGACCTTCAGTGGCCCGAGGAAAAGCTGTACTAGCTGGTGGATATTAGTTTGA	3855
Qy	63	GCCAATTATGATGCTTATGTGACAGCATTTGTCAACGAAATGCATGCAAGTATTAACACC	122
Db	3856	TACAAAATATGAAACATTTGTGATCGGATTTATCGGACGAATGATGCTGAGCCCATCC	3915
Qy	123	AAAGAAGAAC-----CAGTTTGAAGAAGATCTAGATCAAAATTTCTTCACCCAATTGCG	176
Db	3916	TTACGGTTTCATTGCAAGGGGTCTGATTAAGTTTGAAGTGGGTGTGAAGTAACAAATTTAA	3975
Qy	177	AAACGGAATGGGAATATCAATCATCATCAATCAAG--AGAAAGCCAGAGAAGTTCA	233
Db	3976	AGATGGGGAAGTGGCTGATACATATTAATGCTTAAAGTGGCTTCAATCTCTGTTTGATAGG	4035
Qy	234	GTCACGCAATTAATCCATTTTATGAGGCACTATATTCATCGTTTATGCTTAATTAACCC	293
Db	4036	CGGATCTTAAGAACCCCTTTCATTGTAAGAAAGTATCGCTAACGTAATTTAGTACTTTTAAAC	4095
Qy	294	GACCGAAGCATTTGATCTTGAA-----TCATCATTTATCTCAGACCTCGATATCA	344
Db	4096	TAACATGAGACGATCTGCAATTAAGAACTTGTGTATTAATATTTCTCTGATGATGC	4155
Qy	345	TTCAACAAGAAATATCTGAAACCAAGACATCTCGAATGGAAGAAAACATTTCTTTACA	404
Db	4156	CTACACTTCTCAGAGAGATAGCGTTACCGAACATCTGTGCAACAGAAATTAAGATTTTCA	4215
Qy	405	TTCTGTGCGCATTCACGAATGGAAGAACCCGATTAAGTTTCATCGGAGGATTAAGTGC	464
Db	4216	TTCCACACGAATTAAGAAAGTTCCCAAAACAGGGCTGGGCTCTCTCGGAGGTTTAGTAC	4275
Qy	465	AGTTGTGCCCAAGTTTATTTATCCATTTTATCCC-----CAATGTTATCAGTAC	515
Db	4276	AGTTTAACTACACACTTTGGCCCTCTTTTGTATCGGACCTGGAATAATATGTAGCAA	4335
Qy	516	GAATTAAGATATTTTGCACAACGTGCAACAGATTGCAATGTTATGCGCCAAAAAAGAT	575
Db	4336	ATATAGAGAAATATTCATATATTAGCCACAAGTTGCTCATTTGTCAAGCTCAGGGTAAAT	4395
Qy	576	AGGATCTGGTTTATGTTGTCACATGCATGAATTAATGCTGATGTTATATAGAAATTTCA	635
Db	4396	TGGAAGGGGTTTATGATGATGAGCGGGGAGCATATGAACTATCAGATATTAAGAGTTCCC	4455
Qy	636	GCCAGCTTTGATTAATGACGTGTTTCAGGTTCTAGAAAGTGAATCCTGAGAAAGTTCCCAAC	695
Db	4456	ACCCGCAATATCTCTAATTTGCCAGATAT-----GGAAGTCTACTTCGCGCAGTAA	4509
Qy	696	AGAGTTGAAAAAATGTAATGGAAGTAATCGGGAATTCAAACATGAAAGATATACATTACC	755
Db	4510	ACTGGCGATTTGGTGTATGATGAAGAAAGCTGGAATTAATTAACGTAATTAAGTAAACCAATTAACC	4569
Qy	756	ATACGGAATCAAGTTATTAATGAGGTGACGTCAAGGGTGGCTCAGAAAACCCCAATTTGGT	815
Db	4570	TTCCGGATTAACCTTTATGGAATGGGCGAATTTAAGAAATGTTCAAGAAACAGTAAACTGGT	4629
Qy	816	ATCAAGAGTACTCAATGGAAGAAAGAAAGCCAGAAAGAAAGCTCTGTTGTATGACCA	875
Db	4630	CCAGAAAGTAAAAAATGTGATGATTCGCAATATGCGCAAGAAAGCTTGAAAAATATATACAGA	4689
Qy	876	GCTTAATATAGGCCAATTTACAGTTTATGAGAGAAATT	911
Db	4690	ACTGATCTACGAATTTCTAGATTTATGAGATGAGACT	4725

RESULT 21	
AAD31024	
ID	AAD31024 standard, DNA, 7695 BP.
XX	
AC	AAD31024;
XX	
DT	29-AUG-2003 (revised)
DT	31-MAY-2002 (first entry)
XX	
DE	Operon B DNA encoding mevalonate pathway

XX	Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;
KW	isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;
KW	transgenic plant; yeast; phosphomevalonate kinase; HMGR; HMGs; AACT;
KW	mevalonate kinase; mevalonate diphosphate decarboxylase; HMG-CoA; MDD;
KW	acetoacetyl thiolase; 3-hydroxy-3-methylglutaryl-coenzyme A synthase;
KW	MMK; HMG-CoA reductase; ds.
XX	
OS	Saccharomyces cerevisiae.
OS	Arabidopsis thaliana.
OS	Chimeric.
XX	
PN	MO200210398-A2.
XX	
PD	07-FEB-2002.
XX	
PF	31-JUL-2001; 2001WO-US024037.
XX	
PR	31-JUL-2000; 2000US-0221703P.
XX	
PA	(HAHN/) HAHN F M.
XX	(KUEH/) KUEHNLE A R.
PI	Hahn FM, Kuehnle AR;
XX	
DR	WPI; 2002-217122/27.
XX	
PT	Use of specific genes of mevalonate and isoprenoid biosynthetic pathways,
PT	for providing a cell with herbicide or antibiotic resistance, and for
PT	providing transformed cells having increased isoprenoid production.
XX	
PS	Claim 77; Page 132-136; 193pp; English.
XX	
CC	The invention relates to the use of specific genes of the mevalonate and
CC	isoprenoid biosynthetic pathways and inactive gene sites (pseudogene).
CC	Genes of the invention are used to enhance biosynthesis of isopentenyl
CC	diphosphate (IPP), dimethylallyl diphosphate (DMAPP) and isoprenoid
CC	pathway derived products in the plastids of transgenic plants and
CC	microalgae, for producing herbicide or antibiotic resistant transgenic
CC	plants and microalgae, for providing transformed cells with increased
CC	isoprenoid production compared to non-transformed cells, and for
CC	providing a cell with an inserted polynucleotide sequence encoding one or
CC	more products of interest. The present sequence is operon B DNA encoding
CC	the entire mevalonate pathway. This operon contains S. cerevisiae orfs
CC	encoding phosphomevalonate kinase (PMK), mevalonate kinase (MMK),
CC	mevalonate diphosphate decarboxylase (MDD), acetoacetyl thiolase (AACT)
CC	and A. thaliana orfs encoding 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-
CC	-CoA) synthase (HMGs) and HMG-CoA reductase (HMGRC). (Updated on 29-AUG-
CC	2003 to standardise OS field)
XX	
SQ	Sequence 7695 BP; 2214 A; 1524 C; 1793 G; 2164 T; 0 U; 0 Other;
XX	
Query Match	10.2%; Score 132; DB 6; Length 7695;
Best Local Similarity	50.5%; Pred. No. 1.6e-24;
Matches 473; Conservative	0; Mismatches 430; Indels 33; Gaps 5
QY	3 GTCAAAAGCATTTGTGCACTGTGAAAAGCATTTCTTCTGTGGATATTGGTCTTGA 62
DB	28 GTTAGAGCCTTCAGTGCCCAAGGAAAGCGTTACTACTGGTGATATTGTTTGA 87
QY	63 GCCAATTATGATGCTTATGTGACAGCATTTGTTCATCAGATGCATGCAATTAAACAC 122
DB	88 TACAAATATGAAGCATTTGTATGTGCGATTTACGCAAGATGCAATGCTATGACCATCC 147
QY	123 AAAAGAAC-----CAGTTGAAAGATCTAGATCAAAATTTCTTCAACCCCAATTTGC 176
DB	148 TTACGTTTCATTGCAAGGCTCTGATTAATTGAAATGCGGTGAAAGTAACAATTATA 207
QY	177 AAACGGAATGGAATATCATATCATCAATACAG---AGAAGCCAGAGAAGTTCA 233
DB	208 AGATGGGAGTGCGCTATCAATATAGCTCTAAAGTGGCTTCATCTCGTTTGCATAGG 267
QY	234 GTCAAGCATTAATCCATTTTATGAGGCACTATATTCATCGTTTAGCTTATATTCAC 293

```

Db      268 CGGATCTAAGAACCTTTTCATGTAAAAAGTTATCGCTAAGATTTAGCTACTTTAAAC 327
Oy      294 GACCGAAGATTTGATCTTGAA-----TCATCATTTACTAGACCCCTGGATATCA 344
Db      328 TAAACATGAGCACTACTGCAATAGAACTTGTGTATGTATGATTTCTCTGATGATGC 387
Oy      345 TTGACAAGAGATATGAAACCAAGACATCTCGAATGAGAAAAAATTTCTTTACCA 404
Db      388 CTACCAATTTCTCAGAGAGATAGCGTTACCGAACAATCGTGGGACAGAAAGTTGATTTCA 447
Oy      405 TTCTCGTGCATTAACGGAAGTGAAGAACCGGATTAAGTTACGCGCAGATTAAGTGC 464
Db      448 TTGCAAGAAATTTGAAGAGTTCCCAAAACAGGCGTGGGCTCCTCGCGAGTTTAACTAC 507
Oy      465 AGTTGTGCGCAACAAGTTTATATCCCATTTTATCC-----CAATGTTATCAATAC 515
Db      508 AGTTTATACACAGCTTTGGCCCTCTTTTGTATGACCTGGAATAATATGATAGACA 567
Oy      516 GAATTAAGATTTTTCACACAGCTTGCACAGATTTGCAATTTGATGCCCCAAAAAGAT 575
Db      568 ATATAGAGAAGTATATCATATTTAGACAGAGTTGCTCATTTGCAAGCTCAGGGTAAAT 627
Oy      576 AGGATCTGGTTGATTTGATGCACTGCAATTTATGCTGATTTGATTAAGAAATTTCA 635
Db      628 TGAAGCGGGTTGATGAGCGCGGACACATATGATCTATCAGATATAGAAAGTTCCC 687
Oy      636 GCCAGCTTGAATAATGACGTGTTTCAGGTTTCAAGAAATGATCCGGAAGTTCCCCAC 695
Db      688 ACCCGATTAATCTCTAATTTGCCAGATTT-----GGAATGCTACTTACGCGAGTAA 741
Oy      696 AAGATTGAAAAAATGATGAAAGTAACTCTGGAATTCAAACATGAAAGTATGACATTAAC 755
Db      742 ACTGGGCGCTTTGGTTGATGAAAGACGTAATTTACATTAATAAGTAAACCATTTACC 801
Oy      756 ATACGGAATCAAGTATTTAATGAGTGACCTCAAGGTGCTCAGAAACCCCAATTTGCT 815
Db      802 TTCCGATTTAATCTTATGATGAGCGCATATTAAGAAATGTTTGAAGAACGTAATTCGT 861
Oy      816 ATCAGCAGTACTCCATGAGAAAAAGAAAAAGCCAGAAAGAACTGTGTTGTATGACCA 875
Db      862 CCAGAAAGTAAAAAAATGTGATGATTCGATATGCGCAAAAGCTTGAATAATATATACGA 921
Oy      876 GCTTAATAGTCCCATTTACATTTATGAGGAATTT 911
Db      922 ACTCGATATGCAAAATCTAGATTTATGATGATGACT 957

RESULT 22
AAD31027
ID      AAD31027 Standard; DNA; 8224 BP.
XX
XX      AAD31027;
AC      29-AUG-2003 (revised)
XX
XX      31-MAY-2002 (first entry)
DE      Operon E DNA encoding mevalonate pathway and IPP isomerase.
XX
XX      Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;
XX      isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;
XX      transgenic plant; yeast; phosphomevalonate kinase; HMG-R; HMG-S; AACT;
XX      mevalonate kinase; mevalonate diphosphate decarboxylase; HMG-CoA; MDD;
XX      acetoacetyl thiolase; 3-hydroxy-3-methylglutaryl-coenzyme A synthase;
XX      MVK; PMK; HMG-CoA reductase; IPP isomerase; IPP1; ds.
OS      Saccharomyces cerevisiae.
OS      Arabidopsis thaliana.
OS      Streptomyces sp.
OS      Rhodobacter capsulatus.
OS      Chimeric.
XX
XX      WO200210398-A2.

```

```

XX      07-FEB-2002.
PD      31-JUL-2001; 2001WO-US024037.
XX
XX      31-JUL-2001; 2001WO-US024037.
PF
XX
XX      31-JUL-2000; 2000US-0221703P.
PR
XX
XX      (HAIN/) HAIN F M.
PA      (KUEH/) KUEHNIE A R.
XX
XX      Hahn FM, Kuehnle AR;
PI      WPI; 2002-217122/27.
DR
XX
XX      Use of specific genes of mevalonate and isoprenoid biosynthetic pathways,
PT      for providing a cell with herbicide or antibiotic resistance, and for
PT      providing transformed cells having increased isoprenoid production.
XX
XX      Claim 77; Page 145-149; 193pp; English.
PS
XX
XX      The invention relates to the use of specific genes of the mevalonate and
CC      isoprenoid biosynthetic pathways and inactive gene sites (pseudogene).
CC      Genes of the invention are used to enhance biosynthesis of isopentenyl
CC      diphosphate (IPP), dimethylallyl diphosphate (DMAPP) and isoprenoid
CC      pathway derived products in the plastids of transgenic plants and
CC      microalgae, for producing herbicide or antibiotic resistant transgenic
CC      plants and microalgae, for providing transformed cells with increased
CC      isoprenoid production compared to non-transformed cells, and for
CC      providing a cell with an inserted polynucleotide sequence encoding one or
CC      more products of interest. The present sequence is operon E DNA encoding
CC      the entire mevalonate pathway. This operon contains S. cerevisiae orfs
CC      encoding phosphomevalonate kinase (PMK), mevalonate kinase (MVK),
CC      mevalonate diphosphate decarboxylase (MDD), acetoacetyl thiolase (AACT)
CC      and A. thaliana orfs encoding 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-
CC      CoA) synthase (HMS), Streptomyces sp. CL190 orf encoding HMG-CoA
CC      reductase (HMG-R) and R. capsulatus DNA encoding IPP isomerase (IPP1).
CC      (Updated on 29-AUG-2003 to standardise OS field)
XX
XX      Sequence 8224 BP; 2228 A; 1847 C; 2016 G; 2133 T; 0 U; 0 Other;
SQ
XX
XX      Query Match      10.2%; Score 132; DB 6; Length 8224;
XX      Best Local Similarity 50.5%; Pred. No. 1.7e-24;
XX      Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;
Oy      3 GTCAAAAGCATTTTGTGCACTGGAAGAAAGCATTTCTTGGTGGATTTGGTTCTGA 62
Db      43 GTTAGAGCCCTTCACTGTCGCCAGGAAAGCGTTACTGCTGCGATTAATTAAGTTAGA 102
Oy      63 GCCAATTTATGATGCTTATGATGACAGCATTTGTCATCAAGATGATGATGATTAACACC 122
Db      103 TACAATAATATGAAGCATTTTATGATGATGATGATGATGATGATGATGATGATGATG 162
Oy      123 AAAAGAAC-----CAGTTGAAGAAATCTAAGATCAAAATTTCTTACCCCAATTTGC 176
Db      163 TTACGTTTCATTTGCAAGGGTCTGATTAAGTTGAAGTGGTGTGAAAGTAAACAAATTTAA 222
Oy      177 AAACGAGAAATGGGAATATCAATATCATCAATAACAG--AGAAAGCCAGAGAAAGTTCA 233
Db      223 AGATGGGGAATGCTGTCATCAATATTAAGTCTTAAAGTGGTTCATTTCTGTTGATAGG 282
Oy      234 GTACAGCATTAATTCATTTTAAAGCACTATATTAATGATGTTTAAATTCATTAACACC 293
Db      283 CGGATCTAAGAACCTTTTCAATGAAAGTAAAGTTATGCTTAAGTATTAAGTTAAACC 342
Oy      294 GACGGAAGCATTTGATCTTGAAG-----TCATCATTTACTCAGACCCCTGGAATATCA 344
Db      343 TAAACATGACGACTACTGCAATAGAACTTGTTCGTTATTAATTAATTTCTCTGATGATGC 402
Oy      345 TTCAAGAAAGATATGAAACCAAGACATCTCGAATGAGAAAAAATTTCTTTACCA 404
Db      403 CTACATTTCTCAGAGAGATGCGTTACCGAACAATCGTGGACAGAAAGATTTGATTTTCA 462
Oy      405 TTCTCGTGCATTAACGGAAGTGAAGAACCGGATTAAGTTTCAATCGCGAGATTAAGTGC 464

```

Db 463 TTGCGACAGAAATTGAAGAGTTCCCAAAACAGGCGTGGGCTCCTCGGCGGTTAGTCAC 522

Qy 465 AGTTGTTCCACACAGTTATTATATCCCATTTTATCC-----CAATGTTATCAGTAC 515

Db 523 AGTTTAACTACAGCTGGGCTCCCTTTTGTGATCGACCTGGAAAATATATGTACACA 582

Qy 516 GAATTAAGATATTTTGGACACACGTTGCACAGATGCACTTTGTATGCCAAAAAAGAT 575

Db 583 ATATGAGAAAGTTATTCATTAATTATGACACAGTGTCTCATTTGCAAGCTCAGGCTAAAT 642

Qy 576 AGGATCTGGGTTTGATNGTGCACACGCAATTTATGCTGATGTATATGAAAGATTTCA 635

Db 643 TGGAAAGCGGTTTGATGTACGCCGCGCACATATGAGTCTATAGATATATGAAAGATTTCC 702

Qy 636 GCCAGCTTGTATPAAATGACGTGTTTCAGGTTCTAGAAAGTGAATCTGAGAAAGTTCCCAAC 695

Db 703 ACCGCAATTAATCTCTAATTTGCCAGATAT-----GGAAGTCACTTAACGCGAGTAA 756

Qy 696 AGAGTTGAAAAAATGATTGAAAGATACCTGGGATTCAAACATGAAAGATGATCATTACC 755

Db 757 ACTGGCGCAATTTGGTTGATGAAAGAACCTGGATATTAAGATTAAGTAACCATTTACC 816

Qy 756 ATACGGAATCAAGTATTATATGAGGTGACGTCAAGGCTGCTGAAACACCCCAATTTGTT 815

Db 817 TTCGGGATTTAACTTTATGATGATGGCGCATTTTAAGAAATGTTCAAGAAACGTAAATCTGGT 876

Qy 816 ATCAGAGTACTCCAATGGAAGAAAGGAAAGCCAGAGAAAGCTCTGTGTGTATGACCA 875

Db 877 CCAGAAAGGTAAGAAATTTGGTATGATTCGCAATATGCCAGAAAGTTGAAAATATATACAGA 936

Qy 876 GCTTAATATGTGCCAATTTACAGTTATATGAAGATTT 911

Db 937 ACTGCATCATGCAAAATTCAGATTTATGATGATGACT 972

XX	RESULT 23
XX	AAD31025
XX	ID AAD31025 standard, DNA, 8235 BP.
XX	AC AAD31025;
XX	DT 29-AUG-2003 (revised)
XX	DT 31-MAY-2002 (first entry)
XX	DE Operon C encoding mevalonate pathway and IPP isomerase.
XX	XX
XX	Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP
XX	isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;
XX	transgenic plant; yeast; phosphomevalonate kinase; HMGR; HMGs; AACT;
XX	KM mevalonate kinase; mevalonate diphosphate decarboxylase; HMG-CoA; MDD;
XX	acetoacetyl thiolase; 3-hydroxy-3-methylglutaryl-coenzyme A synthase;
XX	KM MKK; PMK; HMG-CoA reductase; IPP isomerase; IPP; ds.
XX	XX
XX	Saccharomyces cerevisiae.
XX	OS Arabidopsis thaliana.
XX	OS Rhodobacter capsulatus.
XX	OS Chimeric.
XX	XX
XX	W0200210398-A2.
XX	XX
XX	07-FEB-2002.
XX	XX
XX	31-JUL-2001; 2001WO-US024037.
XX	XX
XX	31-JUL-2000; 2000US-0221703P.
XX	XX
XX	(HAHN/) HAHN F M.
XX	PA (KUEH/) KUEHNLE A R.
XX	XX
XX	Hahn FM, Kuehnle AR;
XX	XX
XX	WPI; 2002-217122/27.
XX	XX

XX Use of specific genes of mevalonate and isoprenoid biosynthetic pathways
PT for providing a cell with herbicide or antibiotic resistance, and for
PT providing transformed cells having increased isoprenoid production.
XX
XX Claim 77, Page 136-140; 193pp; English.
XX
XX The invention relates to the use of specific genes of the mevalonate and
CC isoprenoid biosynthetic pathways and inactive gene sites (pseudogene).
CC Genes of the invention are used to enhance biosynthesis of isopentenyl
CC diphosphate (IPP), dimethylallyl diphosphate (DMAPP) and isoprenoid
CC pathway derived products in the plastids of transgenic plants and
CC microalgae, for producing herbicide or antibiotic resistant transgenic
CC plants and microalgae, for providing transformed cells with increased
CC isoprenoid production compared to non-transformed cells, and for
CC providing a cell with an inserted polynucleotide sequence encoding one or
CC more products of interest. The present sequence is operon C DNA encoding
CC the entire mevalonate pathway. This operon contains S. cerevisiae orf
CC encoding phosphomevalonate kinase (PMK), mevalonate kinase (MVK),
CC mevalonate diphosphate decarboxylase (MDP), acetosuccinyl thiolase (AACT),
CC A. thaliana orf8 encoding 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-
CC CoA) synthase (HMS), HMG-CoA reductase (HMGR) and R. capsulatus orf
CC encoding IPP isomerase (IPPI). (Updated on 29-AUG-2003 to standardise OS
CC field)
XQ Sequence 8235 BP; 2314 A; 1679 C; 1971 G; 2271 T; 0 U; 0 Other;

Query Match	Similarity	Score	ID	Length
Beat Local	50.5%	Pred. No. 1.7e-24;		
Matches	473;	Conservative	0;	Mismatches 430; Indels 33; Gaps 5
Qy	3	GTCAAAAGCATTTTGTGACCTGTGAAAAGCATTTCTCTGTGTGATATTTGGTCTTGA	62	
Db	28	GTTAGAGCCCTTCAGTGTGCCAGGGAAGCGTTACTACTGTGTGATATTAGTTTAA	87	
Qy	63	GCCAAATTTATGATGCTATGTGACAGCATTTGTATCAGAAATGATGCAATTAAACCC	122	
Db	88	TACAAAATATGAAACATTTGTAGTCGGATTTACGCAAGAAATGCAATGCTGTAGCCATCC	147	
Qy	123	AAAAGAAC-----CAGTTGAAAGATCTAGATCAAAATTTCTTACCCCATTTGC	176	
Db	148	TTAGCGTTTCATTTGCAAGGCTGTGATAGTTTGAAGTGGCTGTGAAAAGTAAACATTTAA	207	
Qy	177	AAACGGAATATGGAATATCATATCATCAATCAAG---AGAAGCCAGAGAAATTCA	233	
Db	208	AGATGGGAGTGGCTGTACATATTAAGCTTAAAGTGGCTTCAATTCCTGTTCCATAG	267	
Qy	234	GTCACGATAAATCCATTTTAAAGGCACTAATATCATCGTTTAACTTATATTCACCC	293	
Db	268	CGGATCTAAGAACCTTTTCATTGAAAAGTTATCGCTAACGTAATTAGCTACTTTTAAAC	327	
Qy	294	GACCGAAGCATTTTATCTTGA---TATCATTTTACTAGACCCCTGGAATATCA	344	
Db	328	TAAATATGACGACTACTGCAATAGAACTTGTCTTATATGATATTTCTGTGATGATGC	387	
Qy	345	TTCAAGAAGATTAAGTGAACCAAGACATCTCTCGAATGGAGAAAACATTTCTTTACA	404	
Db	388	CTACCATTTCTAGAGAGATAGCGTTTACCGAAGTGTGGCAACAGAAATTAAGTTTCA	447	
Qy	405	TTCTCGTGCATTAACGAAAGTGAAGAACCGGATTAAGTTCATGGCAGAAATTAAGTTC	464	
Db	448	TTGCAACGAATTAAGAAAGTTCCCAAAACAGGGCTGGGCTCTCGGCGAGGTTTATGTCAC	507	
Qy	465	AGTTGTGCCCAAGTTTATATCCATTTTATCCC-----CAATGTTATCAATAC	515	
Db	508	AGTTTACTACACCTTTGGCTCTCTTTTGTATCGGACCTGAAAAATTAAGTGACAA	567	
Qy	516	GAATTAAGATTTTGTGCAACGTTGCAAGATTGCAATTTGTATGCGCAAAAAGAAAT	575	
Db	568	ATATAGAGAAATTTCTATTAATTTAGCAACAGTTGCTTATTTGTCAAGCTCAGGGTAAAT	627	
Qy	576	AGGATCTGGGTTTATGTTGCACTGCAATTAATGCTGATATGTATATAGAAAGTTTCA	635	

Db 628 TGAAGCGGTTGATGACGCGGCGACATATGATCTATGATATAGAAATTCCC 687
Qy 636 GCCAGCTTTGATTAATGACGCTGTTTCAGTTCTAGAAAGNATCTGAGAAAGTTCCCCAC 695
Db 688 ACCCGCATTAATCTCTAATTTGCGCAATATTT-----GGAAGTGCCTCTTAACGGCAGTAA 741
Qy 696 AGAGTTGAAAAAATGATGTAAGTAATCTGGGAATTCMAACATGAAAGATGATCAATTACC 755
Db 742 ACTGGCGCATTTGGTATGATGAGAAAGACTGGAATATATAGATTAAAGTAACCATTTACC 801
Qy 756 ATACGGAATCAATTTATTAATGCTGACGTCAGCAAGGCTGCTGAAACCCCAATTGGT 815
Db 802 TTGGGATTAATCTTAATGATGAGCGCATATTAAGAAATGTTTCAGAAACAGTAAACTGGT 861
Qy 816 ATACGAGTACTCCAAATGAAAAAGAAAAAGCCAGAAAGAAAGCTCTGTTGTATGACCA 875
Db 862 CCAGAAAGTAAAAAATTTGGTATGATTCGCAATATGCCAGAAAGCTTAAAAATATATACAA 921
Qy 876 GCTTAATAGTCCAAATTTACAGTTTATGAAAGAAAT 911
Db 922 ACTGATCATGCAAAATTTAGATTATGATGACT 957

RESULT 24

AAD31029

ID AAD31029 standard; DNA; 8400 BP.

XX AAD31029;

AC AAD31029;

DT 29-AUG-2003 (revised)

DT 31-MAY-2002 (first entry)

DE Operon G DNA encoding mevalonate pathway and IDII (IPP isomerase).

XX Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;

KM isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;

XX transgenic plant; yeast; IPP isomerase; IDII, da.

XX Saccharomyces cerevisiae.

OS Arabidopsis thaliana.

OS Schizosaccharomyces pombe.

XX Chimeric.

XX WO200210398-A2.

PN 07-FEB-2002.

XX 31-JUL-2001; 2001WO-US024037.

XX 31-JUL-2000; 2000US-0221703P.

XX (HAHN/) HAHN F M.

PA (KUEH/) KUEHNLE A R.

PI Hahn FM, Kuehnle AR;

XX MPI; 2002-217122/27.

DR Use of specific genes of mevalonate and isoprenoid biosynthetic pathways,

XX for providing a cell with herbicide or antibiotic resistance, and for

PT providing transformed cells having increased isoprenoid production.

XX Claim 77; Page 154-158; 193pp; English.

PS The invention relates to the use of specific genes of the mevalonate and

XX isoprenoid biosynthetic pathways and inactive gene sites (pseudogene).

CC Genes of the invention are used to enhance biosynthesis of isopentenyl

CC diphosphate (IPP), dimethylallyl diphosphate (DMAPP) and isoprenoid

CC pathway derived products in the plasmids of transgenic plants and

CC microalgae, for producing herbicide or antibiotic resistant transgenic

CC plants and microalgae, for providing transformed cells with increased

CC isoprenoid production compared to non-transformed cells, and for

CC providing a cell with an inserted polynucleotide sequence encoding one or

CC more products of interest. The present sequence is operon G encoding

CC entire mevalonate pathway and IDII (IPP isomerase). This operon contains

CC S. cerevisiae, A. thaliana and S. pombe DNA. (updated on 29-AUG-2003 to

XX standardise OS field)

XX Sequence 8400 BP; 2437 A; 1644 C; 1941 G; 2378 T; 0 U; 0 Other;

Query Match 10.2%; Score 132; DB 6; Length 8400;

Best Local Similarity 50.5%; Pred. No. 1.7e-24;

Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

Qy 3 GTCAAAAGCATTTAGTGCACCTGGAAGAAAGCATTTCTGCTGATGATTTGTTCTTGA 62
Db 28 GTTGAAGACCTTCAGTGCCTCCAGGAAAGCCTTACTGCTGTGATTTTGA 87
Qy 63 GCCAATTTATGCTTATGTAAGACAGCTGTCAACGAATGACGATGATTAACACC 122
Db 88 TACAAAATATGAAGCATTTGTAGTCGATATTCGCAAGAAATGATGCTGTAGCCCATCC 147
Qy 123 AAAAGGAAAC-----CAGTTGAAAGAAATCTAGAAATCAAAATTTCTTACCCCAATTGC 176
Db 148 TTACGCTTCAATTCGACAGGCTCTGATTAAGTTGAAGTCGTGAAAGTAAACAAATTTAA 207
Qy 177 AAACGGAATGGGAATATATCATATATCAATATACAG---AGAAGCCAGAGAGATTCA 233
Db 208 AGATGGGAGTGGCTGTACATATATAGTCTTAAAGAGGCTTCAATTCCTGTTGATAGG 267
Qy 234 GTACAGCATTAATCATTTTATGAGGCACTATATTCATGCTTTAGCTTATATCAAC 293
Db 268 CGGATCTAAGAAACCTTTCATTTGAAAAAGTATGCTTAAATTTAGCTTAAAC 327
Qy 294 GACGAAGCATTTGATCTTGAAA-----TCATCATTTTACCAACCCCTGATATCA 344
Db 328 TAAATGAGACATCTACGCAATAGAAACTTGTCTTATGATATTTCTCTGATGATGC 387
Qy 345 TTCAAGAAAGATATCTGAAACCAAGACATCTCGAATGAGAAAAACATTTCTTTTCA 404
Db 388 CTACCATTTCTCAGAGATAGACCTTACCAACATCGTGGCAACGAAGATTTGATTTCA 447
Qy 405 TTCTCGTCCATTACCGAAGTGAAGAACCGATTGTTTCATCGGCAGAGATTAAGTGC 464
Db 448 TTGCAAGAAATTAAGAAAGTTCCCAAAACAGGGCTGGCTCTCGGCAAGTTTAAAGCAC 507
Qy 465 AGTGTTCACCAAGTTTATTTATCCATTTTATCC-----CAATGTTATCACTAC 515
Db 508 AGTTTAACTACAGCTTTGGCTCTTTTGTATGAGACTGGAAAAATATGAGCAA 567
Qy 516 GAATTAAGATATTTTGCACAACTGTCAGATTTGCAATTTGTTATGCCAAAAAAGAT 575
Db 568 ATATGAGAAAGTTATTCATATTTAGCACAAGTTGCTCATTTGTCAGAGGTAAAT 627
Qy 576 AGATCTGGGTTGATGTCGCACTGCAATTTATGCTGATGTTATGAAAGATTTCA 635
Db 628 TGAAGCGGTTTGATGATGAGCGGCGACATATGATCTTATGAAATGAAAGATTTCC 687
Qy 636 GCCAGCTTTGATTAATGACGCTGTTTCAGTTCTAGAAAGTATCTGAAAGTTCCAC 695
Db 688 ACCCGCATTAATCTCTAATTTGCGCAATATTT-----GGAAGTGCCTCTTAACGGCAGTAA 741
Qy 696 AGAGTTGAAAAAATGATGTAAGTAATCTGGGAATTCMAACATGAAAGATGATCAATTACC 755
Db 742 ACTGGCGCATTTGGTATGATGAGAAAGACTGGAATATATAGATTAAAGTAACCATTTACC 801
Qy 756 ATACGGAATCAATTTATTAATGCTGACGTCAGCAAGGCTGCTGAAACCCCAATTGGT 815
Db 802 TTGGGATTAATCTTAATGATGAGCGCATATTAAGAAATGTTTCAGAAACAGTAAACTGGT 861
Qy 816 ATACGAGTACTCCAAATGAAAAAGAAAAAGCCAGAAAGAAAGCTCTGTTGTATGACCA 875
Db 862 CCAGAAAGTAAAAAATTTGGTATGATTCGCAATATGCCAGAAAGCTTAAAAATATATACAA 921
Qy 876 GCTTAATAGTCCAAATTTACAGTTTATGAAAGAAAT 911

XX Saccharomyces cerevisiae.
OS Escherichia coli.
OS Synthetic.
XX US2004005678-A1.
XX 08-JAN-2004.
XX 09-APR-2003; 2003US-00411066.
XX 06-DEC-2001; 2001US-00006909.
XX (KEAS/) KEASLING J.
XX (MART/) MARTIN V.
XX (PITER/) PITERA D.
XX (WITH/) WITHERS S T.
XX (NEWM/) NEWMAN J.
XX Keasling J, Martin V, Pitera D, WITHERS ST, Newman J;
XX WPI: 2004-120864/12.
XX
XX Synthesizing amorpho-4,11-diene in a host cell, useful as
XX pharmaceuticals, comprises introducing nucleic acid sequences, each
XX coding for a different enzyme in the mevalonate pathway for producing
XX isopentenyl pyrophosphate.
XX
XX Disclosure; SEQ ID NO 7; 75bp; English.
XX
XX This invention relates to a novel method of synthesizing amorpho-4,11-
XX diene in a host microorganism which comprises introducing into the host
XX microorganism heterologous nucleic acid sequences, each coding for a
XX different enzyme in the mevalonate pathway for producing isopentenyl
XX pyrophosphate, and introducing into the host microorganism a DNA fragment
XX coding for an optimised synthetic amorpho-4,11-diene synthase gene. The
XX nucleic acid sequences, DNA fragments, vectors and host cells are useful
XX for synthesizing isopentenyl pyrophosphate, amorpho-4,11-diene and epi-
XX cedrol used as pharmaceuticals, nutraceuticals, flavouring agents, and epi-
XX agricultural pest control agents. The present sequence is that of an
XX operon which may be used in the method of the invention.

SO Sequence 9253 BP; 2690 A; 1912 C; 2150 G; 2501 T; 0 U; 0 Other;

Query Match 10.2%; Score 132; DB 12; Length 9253;
Best Local Similarity 50.5%; Pred. No. 1.8e-24;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

3 GTCAAAAGCATTGACCTGGAAGAAAGCATTCTTGTGCTGATATTTGGTTCTTGA 62
6259 GTTGAAGGCTTCAGTCCAGGAGGAAAGCCTTACCTAGCTGGTGAATTTAGTTTGA 6318
63 GCCAATTATGATGCTTATGACAGCATGTCATCAAGATGCATGCAGTTATTAACCC 122
6319 TCAAAATATGAGCATTGTTGATGATGCGCAAGATGCAATGCTGAGCCCATCC 6378
123 AAAAGAAC-----CAGTTGAAAGATCTAGATCAAAATTTCTTCAACCCCAATTGC 176
6379 TTACGTTTCATTCGAAGGCTCTGATAGTTGAATGGCTGTGAAGAAAGTAAACAATTAA 6438
177 AAACGAGAAATGGGAATTCACATTCATCAATACAG--AGAACCCAGAGAAGTTCA 233
6439 AGATGGGAGTGGCTGACCATTAAGTCTTAAAGTGGCTTCACTCTGTTGATAGG 6498
234 GTACAGCATTAATTCATTATTTAGAGGCACTATATTCATGTTAGCTTATATCAAC 293
6499 CGGATCTAAGAACCTTTTCAATGAAAAAGTTTGTGCTAAGATTTAGCTTAAACC 6558
294 GACCGAAGCATTGATCTTGA---TATCATTTTACTCAGAACCTCGGATATCA 344
6559 TAAAGTGAAGCACTAGCAATAGAACTTGTGTTATGATATTTTCTCTGATGATGC 6618
345 TTCAAGAAAGATCTAGAAACCAAGACATCTCGAATGAGAAAAACATTCTTTTACCA 404

DB 6619 CTACATTCCTCAGAGAGATAGGCTTACCGAACATCGTGGCAGAAAGATTGATTTC 6678
QY 405 TTCTCGTCATTTACCGAAGTGAAGAAACCGGATTTAGTTTCATCGGAGATTAGTGC 464
DB 6679 TTGCAAGAAATGGAAGAGTTCACCAAAACGGGCTGGGCTCTCGGAGGTTTAGTCA 6738
QY 465 AGTTGTCACAAAGTTTATTTATCCATTTTATCC-----CAATGTTACAGTAC 515
DB 6739 AGTTTAACTACAGCTTTGGCTCTCTTTTGTATGGAACCTCGAAAAATTAATGACAA 6798
QY 516 GAATTAAGATTTTTCACAAAGTTTCACAGATTGCACTTTATGCCCCAAAAAGAT 575
DB 6799 ATATGAGAAAGTTATTCATTAATTTAGCACAAGTTGCTCATTTGCAAGCTCAGGTA 6858
QY 576 AGCATCTGGGTTTATGATGTCACATGCAATTTATGTCGTGATTTATAGAAATTTCA 635
DB 6859 TGAAGCGGTTTATGATGTCAGGCGGCGCAGCATATGATCTATCAGATATGAAAGTTCC 6918
QY 636 GCCAGCTTTGATTAATGACGTTTCAAGTTTCAAGTTTGAAGAGTTCGAGAAAGTTCC 695
DB 6919 ACCCGCATTAATCTCTAATTTGCCAGATATT-----GGAAGTCTACTTAACGCACT 6972
QY 696 AGAGTTGAAAAAATTTGATTAAGTAAGTAAGTGAATTTCAACATGAAAGATGATCAATACC 755
DB 6973 ACTGCGCATTTGTTGATGTAAGAAAGCTGAAATATTAACGATTAAAGTAACTTTACC 7032
QY 756 ATACGAATCAAGTTATTAATGATGTCAGTCAAGAGTGTGCTCAGAAACCCAAATTTGG 815
DB 7033 TTCCGATTAATCTTATGATAGTGGCGCATTTAAGATGTTTCAAGAAACGATTAATCTGG 7092
QY 816 ATACAGATCTCCTCAAGGAAAGAAAGAAAGCCAGAAAGAGTCTGTTGTATGACCA 875
DB 7093 CCAAGAGTAAAGAAATGATGATGATGTCATATGCAAGAAAGCTTGAATAATATACGA 7152
QY 876 GCTTAATAGTCCAAATTTTACAGTTTATGAAAGAAAT 911
DB 7153 ACTGATCATCAAAATTTCTAATTTATGATGATGACT 7188

RESULT 27

AAD31037 standard; DNA; 13917 BP.

AAD31037;

29-AUG-2003 (revised)

31-MAY-2002 (first entry)

Plasmid transformation vector pK04 DNA.

Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;

isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;

transgenic plant; Yeast; phosphomevalonate kinase; HMGs; ACT;

mevalonate kinase; mevalonate diphosphate decarboxylase; HMG-CoA; MD;

acetate; thiolase; 3-hydroxy-3-methylglutaryl-coenzyme A synthase;

NK; PK; HMG-CoA reductase; ds.

Saccharomyces cerevisiae.

Arabidopsis thaliana.

Chimeric.

MO200210398-A2.

07-FEB-2002.

31-JUL-2001; 2001MO-US024037.

31-JUL-2000; 2000US-0221703P.

(HAHN/) HAHN F M.

(KUEH/) KUEHNLE A R.

PI Hahn FM, Kuehnle AR;
XX WPI: 2002-217122/27.
XX
XX Use of specific genes of mevalonate and isoprenoid biosynthetic pathways,
PT for providing a cell with herbicide or antibiotic resistance, and for
PT providing transformed cells having increased isoprenoid production.
PS
PS Claim 77, Page 161-168; 193pp; English.

The invention relates to the use of specific genes of the mevalonate and
CC isoprenoid biosynthetic pathways and inactive gene sites (pseudogenes).
CC Genes of the invention are used to enhance biosynthesis of isopentenyl
CC diphosphate (IPP), dimethylallyl diphosphate (DMAPP) and isoprenoid
CC pathway derived products in the plastids of transgenic plants and
CC microalgae, for producing herbicide or antibiotic resistant transgenic
CC plants and microalgae, for providing transformed cells with increased
CC isoprenoid production compared to non-transformed cells, and for
CC providing a cell with an inserted polynucleotide sequence encoding one or
CC more products of interest. The present sequence is a plasmid transformation
CC vector pHK04 containing operon B DNA which encodes the entire mevalonate
CC pathway. This operon contains S. cerevisiae orf8 encoding
CC phosphomevalonate kinase (PMK), mevalonate kinase (MVK), mevalonate
CC thalidomide decarboxylase (MDD), acetoacetyl thiolase (AAT) and A.
CC thaliana orf8 encoding 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA)
CC synthase (HMGs) and HMG-CoA reductase (HMGRC). (Updated on 29-AUG-2003 to
CC standardise OS field)

Sequence 13917 BP; 3882 A; 2923 C; 3063 G; 4049 T; 0 U; 0 Other;
SQ

Query Match 10.2%; Score 132; DB 6; Length 13917;
Best Local Similarity 50.5%; Pred. No. 2.1e-24;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

3 GTCAAAAGCATTGATGACCTGGAAGACATTTCTGCTGTGATTTGGTCTTGA 62
4137 GTTGGAGCGCTTCAGTCCCGAGGAAACGTTACTAGCTGGATTTAGTTTGA 4196
63 GCCAATTTATGATGCTTATGACAGCATTTGATCAGAAAGCATGCTATTAACAC 122
4197 TACAAATATGAAAGCTTTGTGTCGATTTATCGGAAAGCATGCTGAGCCATCC 4256
123 AAAAGAAC-----CAGTTGAAAGAACTCTAGAAATCAAAATTTCTTCAACCCCAATTTGC 176
4257 TTAACGTTTCATGCAAGGCTCGATTAAGTTTGAAGCGCTGGAAGAAAGTAAACATTTTA 4316
177 AAACGAGAAATGGAATATCATATTCATCAATAACAG--AGAACCCAGAGAAAGTTCA 233
4317 AGATGGGGAGTGGCTGTACATATTAAGTCTTAAAGTGCTTCATTCCTGTTTCATAG 4376
234 GTACAGCATTAATCCATTTTAGAGGACATATTCATGCTTTAGCTTATATTCAC 293
4377 CGGATCTAAGAACCTTTTCATTAAGAAAGTATTCGCTAAGTATTAAGTACTTTAAAC 4436
294 GACCGAAGCATTTGATCTTGAA-----TCATCATTTTACTCAGACCTGATATCA 344
4437 TAAATGAGACGACTACTGCAATTAAGAACTTTGCTTATGATTTTCTCGATGATGC 4496
345 TTCACAGAAAGTACTGAAACCAAGCATCTCGAATGAGAAAGAAACATTTCTTTACA 404
4497 CTACACATTTCTCAGAGAGATAGGCTTACCGAACATCGTGCAACAGAAAGTATGAGTTTCA 4556
405 TTCTCGTGCATTTACCGAAGTGAAGAAAGCCGATTTAGTTTATCGGCGAGATTAAGTGC 464
4557 TTTCGACAGAAATTTGAAGAAAGTTCACAAACAGGCGTGGGCTCTCGGCGAGTTTATGCTAC 4616
465 AGTTGTGCGACAAAGTTTATATCCATTTTATCC-----CAATGTTATCAGTAC 515
4617 AGTTTATTAACAGCTTGGCTCTTTTGTATGCGACCTGGAAGAAATTAATGTAGACA 4676
516 GAATTAAGATATTTTGCACAGCTTGACAGATTGCAATTTGATGCCCAAAAAAGAT 575
4677 ATATAGAGAAATTTATTAATTTAGCAAGTGTCTCATTTCAAGCTCAGGGTAAAT 4736

QY 576 AGATCTGGGTTTGTATGTTGCAATGCAATTTATGATGTTATATAGAAATTTCA 635
DB 4737 TGGAAAGCGGTTTGTATGACCGCGCAGCATATGATCTATAGAAATTTCC 4796
QY 636 GCGACCTTGTATTAATGACGTGTTTCAGGTTCTAGAAAGTATCTGAGAAATTTCCAC 695
DB 4797 ACCCGCATTAATCTCTAATTTCCAGATATTT-----GGAAGTGTACTTACGCGAGTAA 4850
QY 696 AGATTTGAAATAATGATTTGAAAGTAACTGGAATTTCAACATGAAAGTATTAAC 755
DB 4851 ACTGCGGCTTTGTTGATGAAAGAACTGGAATTTAGATTAAGTAAAGTAACTTAC 4910
QY 756 ATACGAATTCAGTATTATTAATGAGTGTGAGTCAAGGTTGCTCAGAAACCAATTTGCT 815
DB 4911 TTTCGGATTAATCTTTATGATGAGGCGATTTAAGAAATGTTTGAAGAAAGTAACTGCT 4970
QY 816 ATCAGAGTACTCCATGGAAGAAAGAAAGCAGAAAGAACTCTGTTGTATGACCA 875
DB 4971 CCAGAAAGTAAATAATGTTGATGATTCGCAATGCGCAAGAACTTGAATAATATACAGA 5030
QY 876 GCTTAATAGTCCATTTTACAGTTTATGAGAAAT 911
DB 5031 ACTGATCATGCAAAATTTTATGATTTATGATGAGTACT 5066

RESULT 28
AAD31039
ID AAD31039 standard; DNA; 14623 BP.
XX
XX AAD31039;
XX
XX 29-AUG-2003 (revised)
DT 31-MAY-2002 (first entry)
XX
XX DE Plasmid transformation vector pHK08 DNA.
XX
XX Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;
XX isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;
XX transgenic plant; yeast; IPP isomerase; ID11; db.
XX
XX Saccharomyces cerevisiae.
XX Arabidopsis thaliana.
XX Schizosaccharomyces pombe.
XX Chimeric.
XX
XX WO200210398-A2.
XX
XX 07-FEB-2002.
XX
XX 31-JUL-2001; 2001WO-US024037.
XX
XX 31-JUL-2000; 2000US-0221703P.
XX
XX (Hahn/) Hahn F M.
XX (Kuehnle/) Kuehnle A R.
XX
XX Hahn FM, Kuehnle AR;
XX
XX WPI: 2002-217122/27.
XX
XX Use of specific genes of mevalonate and isoprenoid biosynthetic pathways,
PT for providing a cell with herbicide or antibiotic resistance, and for
PT providing transformed cells having increased isoprenoid production.
PS
PS Claim 77, Page 173-181; 193pp; English.

The invention relates to the use of specific genes of the mevalonate and
CC isoprenoid biosynthetic pathways and inactive gene sites (pseudogenes).
CC Genes of the invention are used to enhance biosynthesis of isopentenyl
CC diphosphate (IPP), dimethylallyl diphosphate (DMAPP) and isoprenoid
CC pathway derived products in the plastids of transgenic plants and
CC microalgae, for producing herbicide or antibiotic resistant transgenic

CC plants and microalgae, for providing transformed cells with increased
CC isoprenoid production compared to non-transformed cells, and for
CC providing a cell with an inserted polynucleotide sequence encoding one or
CC more products of interest. The present sequence is plasmid transformation
CC vector pKH06 containing operon G DNA which encodes the entire mevalonate
CC pathway and IDI1 (IPP isomerase). This operon contains S. cerevisiae, A.
CC thaliana and S. pombe DNA. (Updated on 29-AUG-2003 to standardise OS
CC field)

XX Sequence 14623 BP; 4128 A; 3028 C; 3227 G; 4240 T; 0 U; 0 Other;

Query Match 10.2%; Score 132; DB 6; Length 14623;
Best Local Similarity 50.5%; Pred. No. 2.1e-24;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCAAAAGCATTTAGTGCACCTGGAAAAGCATTTCTGTGTGATATTTGGTCTTGA 62
DB 2621 GTTGAGAGCTTCAGTGCCCAAGGAAAGCTTACTAGCTGGGATATTTAGTTTAA 2680
QY 63 GCCAATTTATGATGCTTATGTGACAGCATTTGTCATCAGATGCATGCAATTATACACC 122
DB 2681 TACAAAATATGAAAGCATTTGTAGTCGATTTATCGCAAAAGATGATGCTGATGCCATCC 2740
QY 123 AAAAGGAAAC-----CAGTTGAAAGATCTAGATCAAAATTTCTTCAACCCCAATTGC 176
DB 2741 TTAGGTTTCAATTCAGAGGCTCTGATTAAGTTGAAGTCGTTGAAAAAGTAAACAAATTAA 2800
QY 177 AAAAGGAAATGGGAATATCATATCATCAATATACAG--AGAAGCCAGAGAAAGTTCA 233
DB 2801 AGATGGGAGTGCTGATACATATTAAGTCTTAAAGTGGCTTCTCTGTTTCCATAGG 2860
QY 234 GTACAGCATTAATTCATTTTATGAGCAACTATATTCATGCTTTAGCTTATATTCACCC 293
DB 2861 CGGATCTAAGAAACCTTTTCAATGAAAGATTAAGTCTTAAAGTATTTAGCTTATTAACC 2920
QY 294 GACCGAAGCATTTGATCTTGAAA-----TCATCATTTTCTCAGACCCCGATATCA 344
DB 2921 TAAATGAGAGCACTACTGCATATGAAACTGTTGTTGATTAATTTTCTGATGATGC 2980
QY 345 TTTCACAAAGATATCTGAAACCAAGACATCTCGAATGAGAGAAAAAATTTCTTTACCA 404
DB 2981 CTACCAATCTCAGAGAGATAGCTTACCAACATCGTGCACAGAAATTAAGTTTCA 3040
QY 405 TTCTGTCGATTCACGAGTGGAAAAAGCCGATTTAGTTGATGCGCAGATTAGTGTCC 464
DB 3041 TTTCGACAGAAATGAAAGATTTCCAAAACAGGCGCTCTCTCGGCGAGTTTATGATAC 3100
QY 465 AGTTGTCGACAAAGTTTATATCCCATTTTATCC-----CAATGTTATACAGTAC 515
DB 3101 AGTTTAACTACAGCTTGGCCTCTTTTGTATCGAAGCTGGAAATATATGTAGACAA 3160
QY 516 GAATTAAGATATTTTGCACACGTTGCACAGATTCACATTTGTTATGCCCAAAAAAAGAT 575
DB 3161 ATATGAGAGAGTATATTCATTAATTTAGCAAGTCTCTATGTCAAGCTCAGCGGTAAT 3220
QY 576 AGGATCTGGGTTGATGTTGTCGAACTGCAATTTATGCTGATTTATTAAGAAATTCA 635
DB 3221 TGGAAAGCGGTTGATGAGCGGCGCAGCATATTTGATCTATACATTAAGAAATTTCC 3280
QY 636 GCCAGCTTTGATTAATGACGTTGTTTCAAGTTCTAGAAAGTATCTCGAAGATTTCCAC 695
DB 3281 ACCCGCATTAATCTCTAATTTGCCGATATTT-----GAAAGTCTACTTACGCGAGTAA 3344
QY 696 AGAGTTGAAAAAATGATTTGAAAAGTAACTGGGAATTCAAACATGAAAGATGTACATTAC 755
DB 3345 ACTGCGCATTTGTTGTTGATGAAGAGACTGGAATTTATCAATTAAGAAATTAACATTTAC 3394
QY 756 ATACGGAATCAAGTATTAATGAGGAGTGAAGAGGTGGCTCAAGAAACCAACCAATTTGCT 815
DB 3395 TTGCGGATTAATCTTATGATGAGGCGGATTAATTAAGATGTTTCAAGAAACGTAATTCGT 3454
QY 816 ATCAGAGTACTCCATGAAAAAAGAAAGCAAGAAAGAAAGCTCTGTTGTTATGACCA 875

DB 3455 CCAGAAAGTAAAAAATGATATGATTCGATATGCCAGAAAGCTTGAATAATATACAGA 3514
QY 876 GCTTAATAGTCCATTTACAGTTATGAGAAAT 911
DB 3515 ACTGATATGCAAAATCTAGATTTATGATGACT 3550

RESULT 29
AAD31041
ID AAD31041 standard; DNA; 14623 BP.
XX
XX AAD31041;
AC
XX
XX 29-AUG-2003 (revised)
DT 31-MAY-2002 (first entry)
XX
XX
DE Plasmid transformation vector pKH06 DNA.

KW Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;
KW isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;
KW transgenic plant; yeast; phosphomevalonate kinase; HMGR; HMGs; AACT;
KW mevalonate kinase; mevalonate diphosphate decarboxylase; HMG-CoA; MDD;
KW acetoacetyl thiolase; 3-hydroxy-3-methylglutaryl-coenzyme A synthase;
KW MVK; PKM; HMG-CoA reductase; IPP isomerase; IPPi; ds.

OS Saccharomyces cerevisiae.
OS Arabidopsis thaliana.
OS Streptomyces sp.
OS Rhodobacter capsulatus.
OS Chimeric.

PN W0200210398-A2.
XX
XX
PD 07-FEB-2002.
XX

31-JUL-2001; 2001WO-US024037.
XX
XX
PR 31-JUL-2000; 2000US-0221703P.
XX

(HAHN/) HAHN F M.
PA (KUEH/) KUEHNLE A R.
XX

Hahn FM, Kuehnle AR;
XX
XX WPI; 2002-217122/27.
XX

PT Use of specific genes of mevalonate and isoprenoid biosynthetic pathways,
PT for providing a cell with herbicide or antibiotic resistance, and for
PT providing transformed cells having increased isoprenoid production.

XX Claim 77; Page 185-193; 193pp; English.

XX The invention relates to the use of specific genes of the mevalonate and
XX isoprenoid biosynthetic pathways and inactive gene sites (pseudogene).
XX Genes of the invention are used to enhance biosynthesis of isopentenyl
XX diphosphate (IPP), dimethylallyl diphosphate (DMAPP) and isoprenoid
XX pathway derived products in the plasmids of transgenic plants and
XX microalgae, for producing herbicide or antibiotic resistant transgenic
XX plants and microalgae, for providing transformed cells with increased
XX isoprenoid production compared to non-transformed cells, and for
XX providing a cell with an inserted polynucleotide sequence encoding one or
XX more products of interest. The present sequence is plasmid transformation
XX vector pKH06 containing operon E DNA which encodes the entire mevalonate
XX pathway. This operon contains S. cerevisiae orf encoding
XX phosphomevalonate kinase (PMK), mevalonate kinase (MVK), mevalonate
XX diphosphate decarboxylase (MDD), acetoacetyl thiolase (AACT) and A.
XX thaliana orf encoding 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA)
XX synthase (HMGs); Streptomyces sp. Cl130 orf encoding HMG-CoA reductase
XX (HMGR) and R. capsulatus DNA encoding IPP isomerase (IPPi). (Updated on
XX 29-AUG-2003 to standardise OS field)

Sequence 14623 BP; 4128 A; 3028 C; 3227 G; 4240 T; 0 U; 0 Other;

Query Match 10.2%; Score 132; DB 6; Length 14623;
 Best Local Similarity 50.5%; Pred. No. 2.1e-24;
 Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCAAAAGATTGACCTGGAACAGTATTTCTGCTGGTGAATTTGGTTCTTGA 62
 DB 2621 GTTGAGAGCTTCAGTGGCCCGAGGAAACCGTTACTAGCTGGTGAATTTGGTTTAA 2680
 QY 63 GCCAATTATGATGCTTATGTGACAGATGTCATCAGCAAGTACGAGTATTAACACC 122
 DB 2681 TACAAATATGAAAGCTTTGTAGTCGATTAATCGCAAGAAATGCAATGCTGATCCATCC 2740
 QY 123 AAAGGAAAC-----CAGTTGAAAGAACTAGAAATCAAAATTTCTTCAACCCCAATTGC 176
 DB 2741 TTACGTTTCATGCAAGGCTGTGATAGTTTGAAGTGGCGTGAAGAAAGTAAACAATTAA 2800
 QY 177 AAACGAGATGGGAATATCATATCATCAATACAG---AGAACCCAGAGAGTTCA 233
 DB 2801 AGATGGGGAGTGGCTGTACATATAGTCTTAAAGTGGCTTCAATTCCTGTTTCATAG 2860
 QY 234 GTACAGCATAAATCCATTTTAAAGGCACTATATTATTCATGCTTTAGCTTATATTCAC 293
 DB 2861 CGGATCTAGAAACCTTTTCATTTGAAAGAAAGTATCGTAAAGTATTTAGCTTATTAAC 2920
 QY 294 GACCGAAGCATTTGATCTTGAA-----TCATCATTTTACTCAGACCCTGAGATCA 344
 DB 2921 TAACATGAGCACTACTGCAATATGAAACTTGTTCGTTATTTGATTTCTGATGATGC 2980
 QY 345 TTCACAAAGATATCTGAAACCAAGCATCTCGAATGAGAAAGAAACATTTCTTTACA 404
 DB 2981 CTACCATTTCTCAGAGAGATGAGTATCCAAACATCTGCAACAGAGATGAGTTTCA 3040
 QY 405 TTCTGTCATTAACGAGAGTGAAGAAAGCCGATTAAGTTCAATGGGAGATTAAGTGC 464
 DB 3041 TTGCGACAGAAATGAAAGAGTTCCAAACAGGGCTGGCTCTCGGAGGTTATCTAC 3100
 QY 465 AGTTGTCCACAGATTTATATCCCATTTTATCCC-----CAATGTTATCAGTAC 515
 DB 3101 AGTTTAACTACAGCTTTGGCCCTCTTTTGTATGAGACCTGGAAGATATATGATACAA 3160
 QY 516 GAATTAAGATTTTTCACACGTTGCAACAGATGCAATGTTATGCCCCAAAAAGAT 575
 DB 3161 ATATAGAGAGTATTCATATATTTAGCACAGTGTCTCATTTGCAAGCTCAGGGTAAAT 3220
 QY 576 AGGATCTGGGTTGATGTTGCAACGCAATTTATGCTGATTTATATAGAAATTTCA 635
 DB 3221 TGGAAAGCGGTTGATGAGCGGCGCACATATGATCTATCAGATATAGAAATTTCC 3280
 QY 636 GCCAGCTTGAATTAAGACGTGTTTCAGGTTCTAGAAAGTATCTGAGAAAGTTCCAC 695
 DB 3281 ACCCGATTTATCTCTAATTTGCGCAATAT-----GGAAGCTACTTACGGCAGTAA 3334
 QY 696 AGAGTTGAAAAATGATGAAAGTAACTGGGAATTCAAACATGAAAGATGATACATTAC 755
 DB 3335 ACTGGCGCATTTGGTATGAGAAAGACGTAATATATGACATTAAGATACCATTTAC 3394
 QY 756 ATACGGAATCAAGTATTTATGAGTGAACCTCAAGGTTGCTCAGAAACCCAAATTTGT 815
 DB 3395 TTGGGATTTAACTTTATGAGAGGCGATATTAAGATGTTTCAAGAAACGTAATACCTGT 3454
 QY 816 ATACAGAGTACTCAATGAAAAAGAAAGCCAGAGAAAGCTGTGTTGTATGACCA 875
 DB 3455 CCAGAAAGTAAAAAATGTTGATGATTCGATATGCAAGAAAGTTGAAAAATATATACAA 3514
 QY 876 GCTTATAGTCCCAATTTACAGTTTATGAAAGAAAT 911
 DB 3515 ACTGATCATGCAAAATTTAGATTTATGATGACT 3550

RESULT 30
 ADB69813
 ID ADB69813 standard; DNA; 1506 BP.

AC ADB69813;
 XX 04-DEC-2003 (first entry)
 XX C. neoformans open reading frame SEQ ID NO:2218.
 XX ds; gene; fungicide; gene therapy; infection.
 XX Cryptococcus neoformans.
 XX MO2003052076-A2.
 XX 26-JUN-2003.
 XX 17-DEC-2002; 2002WO-US040225.
 XX 17-DEC-2001; 2001US-0341261P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Zamudio C, Broshkin AM;
 XX WPI; 2003-533017/50.
 XX P-PSDB; ADB70174.
 XX New nucleic acid, useful for preparing a composition for treating an
 XX infection caused by Cryptococcus neoformans.
 XX Claim 2; SEQ ID NO 2218; 136pp; English.
 XX The invention relates to a novel purified or isolated Cryptococcus
 XX neoformans nucleic acid molecule comprising a sequence encoding a
 XX polypeptide comprising a sequence not given in the specification. A
 XX polynucleotide of the invention has fungicide activity, and may have a
 XX use in gene therapy. The nucleic acid is useful for preparing a
 XX composition for treating an infection caused by Cryptococcus neoformans.
 XX The present sequence represents a C. neoformans sequence of the
 XX invention. Note: The sequence data for this patent is not represented in
 XX the printed specification, but was obtained in electronic format directly
 XX from WIPO at ftp.wipo.int/pub/published_pat_sequences.

Sequence 1506 BP; 332 A; 479 C; 342 G; 349 T; 0 U; 4 Other;
 Query Match 5.2%; Score 67.6; DB 10; Length 1506;
 Best Local Similarity 48.8%; Pred. No. 1.6e-07;
 Matches 209; Conservative 0; Mismatches 213; Indels 6; Gaps 1;

QY 718 AGTAATGGGAATTCAAACATGAAAGATGATACATTACATGCAATCAAGTTATATG 777
 DB 1036 AGCAATGGGATATACAAATCTATCCCTTCGACTCCCAACATTTGGCTCTTACTC 1095
 QY 778 GGTACGTCAAGAGGTGGCTCAGAAACACCCAAATTTGTTATCAGATCTCCAAATGAA 837
 DB 1096 GCCGACGTTTCGTGGGTACTGACACGCGCATCTTGTCTCTCGTCTCAAGTGA 1155
 QY 838 AAGGAAAGCCAGAGAAAGCTCTGTGTTGATGACAGCTTAATAGTCCAAATTACAG 897
 DB 1156 AATTAACATAGGAAAGAGCCGATAGTGTGGGAGGCTGATATATGCAACAGGGC 1215
 QY 898 TTATGAAGAAATGAGGAAATGCGTGAATAATAGACTAGACCCAGAGACTTATAT 957
 DB 1216 TTGGGTAGGTGCTGAAAGATATGCTTTGAACCTCCCTACAAACCCGACCTTCAGC 1275
 QY 958 AAAGATTAATCATTTCTGTAGAGCTTTGACGTGTGGATTAAGAAATCAGAAAAAGG 1017
 DB 1276 ACATCT-----CATNTCTCCACCGCCCTGCTNTCCCTTNTGTCATCGCGCTGTG 1329
 QY 1018 TTACAAGCATTAACAACAAATATCAGAGTTCCAAATGAACTGATGCCAACCAAGTTG 1077
 DB 1330 TTGCGGAAATGTCCGACTCTCGGCGTACCATTCGAAACAAAGAACCAAGATTG 1389
 QY 1078 TTGAACCGTTGCAAGAGATTCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGA 1137

Dd	1390	CTGATGCATCGGACAAAGTCAGAGGTGTGAGCGGTGGCGTTCAAGTCCGAGCCGC	1449
QY	1138	TACGATGC	1145
Dd	1450	TACGATGC	1457

RESULT 31

ID ABT19236 standard; DNA; 1356 BP.

AC ABT19236;

DT 16-APR-2003 (first entry)

DE Aspergillus fumigatus essential gene #1594.

XX
KW
LM
fungicides; cyclostatic; essential gene; *Aspergillus fumigatus*; infection;
cancer; contamination; biofilm; antibody; immune response; ds.

Aspergillus fumigatus.

PN WO200286090-A2.

PD 31-OCT-2002.

PF 23-APR-2002; 2002WO-US013142.

23-APR-2001; 2001US-0285697P.

05-JUN-2001; 2001US-0295890P.

PR 31-AUG-2001; 2001US-0316362P.
xx

PA (ELI)- ELITRA PHARM INC.
XX

Ulang B, Tishkoff D, Zamudi

WPL; 2003-093124/08.

PT New purified or isolated nucleic acids of essential genes of *Asperg*
fumigatus useful for

or for creating a new
 F1
 XX

Disclosure; Page: 175pp; English.

essential genes of *Acetivibrio* 117

The invention are used to treat or prevent infections by a pathogenic organism such as *A. fumigatus*, to treat a non-infectious disease in a subject (e.g. cancer), to prevent or contain contamination of an object by *A. fumigatus*, or to prevent or inhibit formation on a surface of a biofilm comprising *A. fumigatus*. The polynucleotides are useful for expressing recombinant protein for characterisation, screening or therapeutic use, as markers for host tissues in which the pathogenic organisms invade or reside, for comparing with the DNA sequence of *A. fumigatus* to identify duplicated genes or paralogues having the same or similar biochemical activity and/or function, for comparing with DNA sequences of other related or distant pathogenic organisms to identify potential orthologues essential or virulence genes, for selecting and making oligomers for attachment to a nucleic acid array for examination of expression patterns, for raising anti-protein antibodies, as an antigen to raise anti-DNA antibodies or to elicit another immune response, and for identifying polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction. The polypeptides may be used to raise antibodies or to elicit immune response, as a reagent in assays designed to quantitatively determine levels of the protein in biological fluids, as a marker for host tissues in which pathogenic organism invade or reside, and to isolate correlative receptors or ligands in the case of virulence factors. This polynucleotide sequence represents one of the essential genes of *Aspergillus fumigatus* of the invention

SQ Sequence 1356 BP; 293 A; 384 C; 384 G; 295 T; 0 U; 0 Other;

Query Match	Score	DB	Length
5.18;	66;	8;	1356;

Matches 547; Conservative 0; Mismatches 595; Indels

6 AAAGCATTTAGTCACCTGGAAGCATTTCTGCTGGTGATATTTGGTCTTGAGCC 65

6 AAAAGCATTTAGTGCACCTGGAAAAAGCATTTCTTCGTCGTGTGATATTTGGTCTTTGAGCC 65
24 AAGCGCTTGTCCGCGCCGGGCAAGGTCTCTCACTGGGGATATCTGTGCTCGAGCCG 83
66 AATTATGATGCTTATGTGACAGCAATTGTCAATCAGAAATGCATGCATTATACCAAAA 125
84 CACCTCACCTGGGACGTGTGTGCCCTTGAAGCCAGAAATCATGTGGAAAGGCANATTTCA 143
126 AGGACACGAGTTTGAAGAATCTAGATATCAAAATTTCTTACCCTCAATTTGCAACGAGA 185
144 TGGAGCAGAAGAGCAAGAGGACGATGTGTATACATCTCCACAGTTCGGGATGGGT 203
186 ATGGGAATATCATATATCATCAATATACAGAAAGCCAGAAAGTTGACATGCATATAA 243
204 ATGGGAGTATACATATACAAAGATGGGAGACGAGCTGACCAATCCAGGCGCAGCC---AA 266
246 TCCATTTTAAAGGGAACATATATCAATGTTTAGCTTATAT-----TCAACGAC 297
261 CCTGTTTTCGAGACCTCTCTGAATTTGCCCTTACATTAATCAGTATGTGGCCGATC 320
298 GAAGCATTTGATCTTGAATATCATATTTACTCAGACCTGGATATCATTCACAAGAAAT 357
321 GAAGATTTCCGGGTATATGATATGATATCTCTCGCCGACAAAGATTAATCAATCCAGAC 380
358 ACTGAACCAAGACATCTCTGAATGGAG-----AAAAACATTTCTTACCATTCTGTGC 413
381 GGCCTTCTCAGAGCTTGGAGCTTCGGTGTCTCAGAGATTCGTGTGACTTGGATGTGC 440
414 CATTAACGAGTGGAAAAAGACCGGATTAGTTCAATCGGACAGATTAGTGTCACTTGTGC 473
441 CCTTCAGAGGACACAAGAGCAGGCTGGGGCTCTTCAAGCCGCTGTGTCACTGCCCTGCT 500
474 CACAACTTATTTATCCATTTTA-----TCCCAATGTTATCATGATACGAATTAAGATAT 527
501 GTCTGCTCTGTGTCAATCCACGGTACTATGCAACCCAGATCTGTGTCTCAGGCCACACAA 560
528 TTTCACACACGTTGCACAGATTCGACATTTGTATATGCCCCAAAAAAGATAGGATCTGGATT 587
561 GCTTCAACATCTGGCCCCAGGGGGCCACTGGCTGGCGAGGGTAATGGGGTCCGGCTT 620
588 TGATGTTGCAATCTGCATTTTATATGTCTGATTTGATATAGAGATT---TCAGCCACTT 643
621 CGATGTGCACTGCAATTTACGGCTCTGTCTCTACAGACAGCGTTCTTCCCTCTGATTTCT 680
644 TGATAAATGACGTGTTTCAGGTTCTAGAAAGTATCTGAGAAAGTTCCCCACAGAGTTGA 703
681 CGAATCAGTGGGTGACCGCTGTTCTCCAGGCTTCCGAAGCCGTTCTTCCGATCGTAGA 740
704 AAAAATTGATTTGAAGAT-----ACTGGGAATTCAAAATGAAAGATGTACATTACC 755
741 GGAAGCCGACCCCGCAGATCTGTGGACACACGAAATGTACACTTCGGAGTAACCTCC 800
756 ATACGGAATCAATTTATTTAATGGGTACGTCAGAGGTGCTCAGAAAACCCCAATTTGTT 815
801 CCGGGGAATCAATGTGCTCTGTACGTTCGAATGTGTTTCGACATCTCATTCATGTT 860
816 ATCAACGATACTCATATGAAAAAAGAAAAAGCAGAAAGAAAGCTCTGTGTGTATGACCA 875
861 GAGAAAGTTTTGAGAGTGGGAGAAACAGATCAGAAAGAAAGCGAATATGCTCTGGGGCGC 920
876 GCTTAATAGTCAATTTACAGTTTATGAAAGAAATGAGGGAATAGCGTGAATAAATAAGA 935
921 TCTGCAATCGAACAAGAGAGACTTCGCTGAAATCTCAGACGCTTGCACAGAGCCGGA 980
936 CTCAAGCCACAGACCTTATATTTAAAGATTTAGATCAATCTCTGTGAGCCTTGTGACTGTTC 995
981 CGAACATATCTCAGTGAATTTGAAAAATCCGACACTAATATTCAGCCTG----- 10322

QY 996 GATTAGAAATCAGAAAGGCTTACACGATTAACACAAATACAGAGTTCCATTGA 1055
 DB 1033 -----CGTAACACATCCGTTCCATGACTCAAAAGTCGATGCCAATCA 1079
 QY 1056 ACCGTATGTCACACCCAGTTTGGACCGTGTCAAGATTCCTGTTGTTGTGTG 1115
 DB 1080 GCCGGCGCTCAACCCAGCTACTGACCTGTCCGAGCTGAGAGGTGCATCGTGG 1139
 QY 1116 TGTGTTCCAGGTGTGTGATACGATCAATAGCTGATTAGTGTGAAATC 1171
 DB 1140 TGTGTTCCAGAGCAGGGGGCTACGACCATTTGCTCTCATCCAGACATC 1195
 RESULT 32
 AB228940/C
 ID AB228940 standard; DNA; 65 BP.
 XX AB228940;
 AC
 XX 30-JAN-2003 (first entry)
 XX
 DE Candida gene related tetracycline promoter PCR primer SEQ ID NO 3023.
 XX
 KM Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
 KM signal transduction; DNA replication; cell division; growth;
 KM proliferation; Candida albicans; fungicide; antifungal; PCR; primer; ss.
 OS
 XX Candida albicans.
 XX
 XX WO200253728-A2.
 XX
 XX 11-JUL-2002.
 XX
 PD 26-DEC-2001; 2001WO-US049486.
 PF
 XX 29-DEC-2000; 2000US-0259128P.
 PR 20-FEB-2001; 2001US-00792024.
 PR 22-AUG-2001; 2001US-0314050P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Roemer T, Jiang B, Boone C, Buesey H, Ohlsen KL;
 DR WPI; 2002-56694/60.
 XX
 PT Constructing strains for identifying gene products as effective targets
 PT for therapeutic intervention, by inactivating in the strain one allele of
 PT a gene and placing other allele of the gene under conditional expression.
 XX
 PS Claim 76; SEQ ID NO 3023; 167pp + Sequence Listing; English.
 XX
 CC The invention relates to constructing (M1) a strain of diploid fungal
 CC cells in which both alleles of a gene are modified, comprising modifying
 CC one allele by insertion or replacement by a cassette having an
 CC expressible selectable marker and modifying another allele by
 CC recombination, of a promoter replacement fragment with a heterologous
 CC promoter, so that expression of the second allele is regulated by the
 CC promoter. (M1) is useful for constructing a strain of diploid fungal
 CC cells in which both alleles of a gene are modified. The diploid fungal
 CC cells having both alleles modified are useful for identifying a gene that
 CC is essential to the survival or growth of a fungus, a gene that
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene
 CC that contributes to the resistance of a diploid fungus to an antifungal
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus
 CC and for identifying a therapeutic agent for treatment of a mammalian
 CC disease. (M1) is useful for identifying a compound which modulates the
 CC activity of a gene product, preferably enzymatic activity, carbon
 CC compound catabolism, biosynthesis, transporter, transcriptional,
 CC translational, signal transduction, DNA replication and cell division
 CC activity. The method is useful for identifying a compound having the
 CC ability to inhibit growth or proliferation of C. albicans cells and for
 CC treating infection by C. albicans. The present sequence is that of a PCR

CC primer used in the method of the invention. Note: The sequence data for
 CC this patent is not represented in the printed specification but is based
 CC on sequence information supplied to Derwent by the European Patent Office
 SQ Sequence 65 BP; 22 A; 16 C; 11 G; 16 T; 0 U; 0 Other;
 Query Match 5.0%; Score 65; DB 6; Length 65;
 Best Local Similarity 100.0%; Pred. No. 2.2e-07;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGTAAAGCATTTAGTGCACCTGAAAGCATTTCTTGTGATTTGTTCTT 60
 DB 65 ATGTAAAGCATTTAGTGCACCTGAAAGCATTTCTTGTGATTTGTTCTT 6
 QY 61 GAGCC 65
 DB 5 GAGCC 1
 RESULT 33
 AAS23667/C
 ID AAS23667 standard; DNA; 90 BP.
 XX AAS23667;
 AC
 XX 04-DEC-2001 (first entry)
 XX
 DE Tetracycline promoter downstream PCR primer (Tet-Down) #42.
 XX
 KM Gene identification; essential gene; GRACE; pathogenic fungus;
 KM gene replacement and conditional expression; fungal infection;
 KM PCR primer; Tet-Down; tetracycline promoter; ss.
 OS
 XX Escherichia coli.
 XX
 XX Candida albicans.
 XX
 XX WO200160975-A2.
 XX
 PD 23-AUG-2001.
 PF
 XX 20-FEB-2001; 2001WO-US005551.
 PR 18-FEB-2000; 2000US-0183534P.
 PR
 XX (ELIT-) ELITRA PHARM INC.
 XX
 PI Roemer T, Jiang B, Boone C, Buesey H;
 DR WPI; 2001-489080/53.
 XX
 PT Identifying genes essential to fungal metabolisms and identifying
 PT potential therapeutic agents that target these genes.
 XX
 PS Disclosure; Page 297; 324pp; English.
 XX
 CC The present invention relates to novel methods for constructing fungal
 CC strains useful for identification and validation of gene products as
 CC targets for therapeutic agents, for creating a collection of identified
 CC essential genes, and screening assays for the discovery of new drugs. The
 CC invention provides the GRACE (gene replacement and conditional
 CC expression) method for the construction of mutant organisms referred to
 CC as GRACE strains of the organism. The invention can be applied to any
 CC organism, particularly a pathogenic fungus e.g. Candida albicans,
 CC Aspergillus fumigatus and Cryptococcus neoformans. The methods are useful
 CC to identify agents that may be used in the treatment of fungal
 CC infections. AAS23626-AAS23686 represent tetracycline promoter downstream
 CC PCR primers (Tet-Down) used in the methods of the present invention
 SQ Sequence 90 BP; 28 A; 19 C; 18 G; 25 T; 0 U; 0 Other;
 Query Match 5.0%; Score 65; DB 4; Length 90;
 Best Local Similarity 100.0%; Pred. No. 2.5e-07;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANGCAAAAGCATTTAGTGAACCTGGAAGCATTTCTGTGTGATATTTGTTCTT 60
DB 65 ATGTAAAAAGCATTTAGTGAACCTGGAAGCATTTCTGTGTGATATTTGTTCTT 6
OY 61 GAGCC 65
DB 5 GAGCC 1

RESULT 34
ABT21056
ID ABT21056 standard; DNA; 1455 BP.
AC ABT21056;
XX
XX
DT 16-APR-2003 (first entry)
XX
XX Aspergillus fumigatus essential gene #3414.
DE
XX Aspergillus fumigatus essential gene; Aspergillus fumigatus; infection;
KW Fungicide; cytosolic; essential gene; Aspergillus fumigatus; infection;
XX cancer; contamination; biofilm; antibody; immune response; de.
OS Aspergillus fumigatus.
XX
XX WO200286090-A2.
XX
XX 31-OCT-2002.
XX
XX 23-APR-2002; 2002MO-US013142.
XX
XX 23-APR-2001; 2001US-0285697P.
XX 27-APR-2001; 2001US-0287066P.
XX 05-JUN-2001; 2001US-0295890P.
XX 09-JUL-2001; 2001US-0303899P.
XX 31-NOV-2001; 2001US-0316362P.
XX
XX (ELITRA PHARM INC.
XX
XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX WPI; 2003-093124/08.
XX
XX New purified or isolated nucleic acids of essential genes of Aspergillus
XX fumigatus, useful for treating or preventing infections by A. fumigatus,
XX or for treating a non-infectious disease in a subject e.g. cancer.
XX
XX Disclosure; Page; 175bp; English.

XX
XX The invention relates to novel purified or isolated nucleic acids of
XX essential genes of Aspergillus fumigatus. The isolated nucleic acids of
XX the invention are used to treat or prevent infections by a pathogenic
XX organism such as A. fumigatus, to treat a non-infectious disease in a
XX subject (e.g. cancer), to prevent or contain contamination of an object
XX by A. fumigatus, or to prevent or inhibit formation on a surface of a
XX biofilm comprising A. fumigatus. The polynucleotides are useful for
XX expressing recombinant protein for characterization, screening or
XX therapeutic use, as markers for host tissues in which the pathogenic
XX organisms invade or reside, for comparing with the DNA sequence of A.
XX fumigatus to identify duplicated genes or paralogues having the same or
XX similar biochemical activity and/or function, for comparing with DNA
XX sequences of other related or distant pathogenic organisms to identify
XX potential orthologous essential or virulence genes, for selecting and
XX making oligomers for attachment to a nucleic acid array for examination
XX of expression patterns, for raising anti-protein antibodies, as an
XX antigen to raise anti-DNA antibodies or to elicit another immune
XX response, and for identifying polynucleotides encoding the other protein
XX with which binding occurs or to identify inhibitors of the binding
XX interaction. The polypeptides may be used to raise antibodies or to
XX elicit immune response, as a reagent in assays designed to quantitatively
XX determine levels of the protein in biological fluids, as a marker for
XX host tissues in which pathogenic organisms invade or reside, and to
XX isolate correlative receptors or ligands in the case of virulence

CC factors. This polynucleotide sequence represents one of the essential
CC genes of Aspergillus fumigatus of the invention
XX
XX Sequence 1455 BP; 318 A; 406 C; 419 G; 312 T; 0 U; 0 Other;
SQ

Query Match 4.8%; Score 62.6; DB 8; Length 1455;
Best Local Similarity 46.7%; Pred. No. 3.4e-06;
Matches 375; Conservative 0; Mismatches 389; Indels 39; Gaps 4;

OY 387 AAAAAGCATTTCTTACCATTCCTGCGCATTAACCGAAGTGAAGAAAGCCGATTAGTTTC 446
DB 513 AAGCATTTCTGATGACCTTGTGTGCTTCAGAGGACACAAAGAGCCGCTGCTC 572
OY 447 ATCGCAGATTAAGTGTGATGTTGCCAAGTTTATATCCCA-----TTTATCCC 500
DB 573 TTCAAGCCGCTTGGTCACTGCGCTGTGTGTCTCTCTGTCATCAACCGTACTATGCAAC 632
OY 501 CAATGTTATCAATGATGAATTAAGATATTTTGCACAACTTGCACAGATTGCAATGTTA 560
DB 633 CGACGATCTCGGTCCAGGCGCGGACCAAGCTTCAATGAGCCCGGCGCCACTGCGC 692
OY 561 TGCCCAAAAAAGATAGATCTGGGTTGATGTTGCACTGCAATTATATGTTGATGTTGT 620
DB 693 TGCCGAGGTTAAAGTCCGCTCCGCTCCGCTTGAATGTTGAGCTGCATTAAGGCTCCT 752
OY 621 ATATGAAAGATTTCAGCCAGCTTTGATTAATGACGTGTTCA---GGTTCTGAAAGTG 676
DB 753 CTACAGAGCTTCTCCCTCGATTCGATTCAGATGAGGTGACGCTGCTTCCAGGCTT 812
OY 677 ATCTGAAAGATTCCTCCCAAGATGGAAGAAATTTGAAAGTAACTGGG-----A 728
DB 813 CGAAGAGCGGTTGTTCCGATGCTGAGAGAGCGCCAGCCCGACATCCGTGGAGACCGA 872
OY 729 ATTCAACATGAAGAATGATGATTAATCAATGAGATCAAGTATTAATGAGTGAACGTGA 788
DB 873 ATGTTAATGATTCGAGATGAGAGCTCCCGCGGAAATGAAATGCTCTGTGAGCTGGA 932
OY 789 GGGTGGCTCAGAAACACCCCAATGATGATCAAGATCTCAATGAGAAAGGAAAGCC 848
DB 933 ATGTGTTTCGAGACTCATCATGATGAGAAAGTTTGAAGTGGCGGAAACGAAATCA 992
OY 849 AGAAGAAAGCTCTGTGTGATGATGACCAAGTAAATGATTAAGTTATGAAGCA 908
DB 993 GAAAGAAACCGATATGCTCTGGGCGCTCTGATGATGAAACGAGAGCTTCGCTGGA 1052
OY 909 ATTGAGGAAATGCTGAGAAATACGATCAGACCCAGAGACTTATATTAAGATTGA 968
DB 1053 ACTCAGAGCTTGGACAGAGCCGGAAGACATATCTCAGTGACTTTGAAATGTCGG 1112
OY 969 TCATTCGTGTGAGCTTTGACTGTTGCGATTAAGAACATCGAAGAGGTTTCAAGCAT 1028
DB 1113 CACATATATTCAGCGCTCG-----CGTAAACCATCCGTTCCAT 1151
OY 1029 AACCAAAATACAGAGTTTCAATTAAGCTGATGTCGAACCCGATTGTTGACCGTTG 1088
DB 1152 GACTCAAAAGTCGATGATTCATATGAGCGCGCTTCAACCGAGCTATTTGAGCTCT 1211
OY 1089 TCAAGAGATCTCTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1148
DB 1212 GTCCAGAGCTGAGAGGTTGATCGATGATGTTGTTGTTGTTGTTGTTGTTGTTGTT 1271
OY 1149 AGCTGATTAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1171
DB 1272 TGTGCTCTCATCAAGACAAATC 1294

RESULT 35
ABT20458
ID ABT20458 standard; DNA; 1509 BP.
XX
XX
AC ABT20458;
XX
XX 16-APR-2003 (first entry)
DT

XX Aspergillus fumigatus essential gene #2816.
 DE Fungicide; cytosstatic; essential gene; Aspergillus fumigatus; infection;
 XX cancer; contamination; biofilm; antibody; immune response; de.
 KW Aspergillus fumigatus.
 XX
 XX WO200286090-A2.
 XX
 PD 31-OCT-2002.
 XX
 XX 23-APR-2002; 2002WO-US013142.
 XX
 XX 23-APR-2001; 2001US-0285697P.
 PR 27-APR-2001; 2001US-0287066P.
 PR 05-JUN-2001; 2001US-0295890P.
 PR 09-JUL-2001; 2001US-0303899P.
 PR 31-AUG-2001; 2001US-0316362P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX
 XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM,
 DR WPI, 2003-093124/08.
 XX
 PT New purified or isolated nucleic acids of essential genes of Aspergillus
 PT fumigatus, useful for treating or preventing infections by A. fumigatus,
 PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX
 XX Disclosure; Page; 175pp; English.
 XX
 CC The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as A. fumigatus, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for
 CC expressing recombinant protein for characterization, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of A.
 CC fumigatus to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of Aspergillus fumigatus of the invention
 XX
 XX Sequence 1509 BP; 338 A; 418 C; 429 G; 324 T; 0 U; 0 Other;
 XX
 XX Query Match 4.8%; Score 62.6; DB 8; Length 1509;
 XX Best Local Similarity 46.7%; Pred. No. 3.5e-06;
 XX Matches 375; Conservative 0; Mismatches 389; Indels 39; Gaps 4;
 QY 387 AAAAATTTCTTACATCTGTCGATTCAGGAGGAGAAAGACCGGATTAGTTC 446
 DB 567 AAGCAGATTGCGTGAATTTGTTGCTTCAAGAGGACACAAAGAGGCTGGCTC 626
 QY 447 ATCGCAGATTAGTGTGATGTTGCTTCAAGGAGGAGGAGGAGGAGGAGGAGG 500
 DB 627 TTCAGCGCGCTTGTGATCAGCCCTGCTGCTCTCTGCTGCTGCTGCTGCTGCTG 686

QY 501 CAATGTATCAATGAAATTAAGATATTTTGCACAACTGTCACAGATTGCAATTTGTA 560
 DB 687 GCAAGATCTCGGTCCAGGCGCGGACAGCTTCAATCTGGCCAGGCGGCCACTGCGC 746
 QY 561 TGCCCAAAAAAAGATAGATCTGGTGTGATTTGCACTGCAATTTATGCTGATTT 620
 DB 747 TGCCGAGGATAAAGTCGGCTCGGCTGATGTTGACCTGCAATTTACGGCTCTGCT 806
 QY 621 ATATGAAATTTTCAAGCAGCTTTGATTAATGACGTTTCA-----GGTTAGAAAGTG 676
 DB 807 CTACGAGCGCTTCTCCCTCGATTTCTGATCTGATGAGGTGACGCTGTTCTCCAGGCT 866
 QY 677 ATCTGAGAAAGTTTCCCAAGAGTTGAAATTTGAAATGAAATGAAATGAAATGAA 728
 DB 867 GGAAGAGGAGTTGTTCCGAGATGTAAGAGAGCCGACCCCGAGCATCCGTGGACACGA 926
 QY 729 ATTCAAAATGAAATGATGATCTTCAATGAGGATCAAGATTTAATGAGTGAAGTCA 788
 DB 927 ATGTCTAAGCTTTCGAGATGAAAGCTCCCGGAGATGCAAAATGATGTCCTGTGAGCTGA 986
 QY 789 GGGTGGCTCAGAAACACCCAAATTTGTTATCAGAGTCTCAATGAGAAAGAAAGGCG 848
 DB 987 ATGTGTTTCCAGACTTCCATTCATGTAAGAAAGTTTGGAGTGGCGGAAACAGATTA 1046
 QY 849 AGAAGAAAGCTCTGTGTGTATGACCACTTAATAGTCCAAATTTACAGTTTATGAGA 908
 DB 1047 GAAAGAGCGCATATGCTCTGGGCGCTGTGCAATGCAACAGAGACTTCGCTGGA 1106
 QY 909 ATTGAGGAAATGCGTGAATTAATGACTCAGACCCAGAGCTTATTTAAAGAGTTAGA 968
 DB 1107 ACTCAGAGCGTTGGACAGAGCCCGAGCAATCTCTGATGACTTTGAAATGTCG 1166
 QY 969 TCATCTGTGAGCTTTGACGTGTTGCGATTAAAGATCAGAAAGGTTACAGACTT 1028
 DB 1167 CACCTATATCAAGCCTCG-----CTTAACACATCGTTTCAAT 1205
 QY 1029 AACCAAAAATCAGAGTTTCCATTAATGAACCTGATGCCAACCAGTTTGAACCGTGG 1088
 DB 1206 GACTCAAAAAGTGGATGTCCTCAATGAGCGCGGCTCCAAACGAGTACTTGAAGCTCT 1265
 QY 1089 TCAAGAGATCTCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1148
 DB 1266 GTCCAGCGTGAAGGATGTCATCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1325
 QY 1149 AGCTGATTAAGTGTGAATTC 1171
 DB 1326 TGTGCTCTCATCAAGCAATC 1348
 RESULT 36
 ID ABT18642
 ID ABT18642 standard; DNA; 1509 BP.
 XX ABT18642;
 AC 16-APR-2003 (first entry)
 XX
 DT Aspergillus fumigatus essential gene #1000.
 XX
 XX Fungicide; cytosstatic; essential gene; Aspergillus fumigatus; infection;
 KW cancer; contamination; biofilm; antibody; immune response; de.
 XX
 XX Aspergillus fumigatus.
 XX
 XX WO200286090-A2.
 XX
 PD 31-OCT-2002.
 XX
 XX 23-APR-2002; 2002WO-US013142.
 XX
 XX 23-APR-2001; 2001US-0285697P.
 PR 27-APR-2001; 2001US-0287066P.
 PR 05-JUN-2001; 2001US-0295890P.

PR 09-JUL-2001; 2001US-0303899P.
 PR 31-AUG-2001; 2001US-0316362P.
 PA (ELIT-) ELITRA PHARM INC.
 PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 DR WPI; 2003-093124/08.
 XX New purified or isolated nucleic acids of essential genes of *Aspergillus*
 PT fumigatus, useful for treating or preventing infections by *A. fumigatus*,
 PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX Disclosure; Page: 175pp; English.
 PS
 XX The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as *A. fumigatus*, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or inhibit formation of an object
 CC by *A. fumigatus*, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising *A. fumigatus*. The polynucleotides are useful for
 CC expressing recombinant protein for characterisation, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of *A.*
 CC *fumigatus* to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of *Aspergillus fumigatus* of the invention
 XX
 SQ Sequence 1509 BP; 338 A; 418 C; 429 G; 324 T; 0 U; 0 Other;
 Query Match 4.8%; Score 62.6; DB 8; Length 1509;
 Best Local Similarity 46.7%; Pred. No. 3.5e-06;
 Matches 375; Conservative 0; Mismatches 389; Indels 39; Gaps 4;
 QY 387 AAAAAATTCTTACCATCTCGTGCCATTACGAGTGAAGAAACGCGATTAGGTT 446
 DB 567 AACGACATTCGTGGAATTGTTGTCGCTTCAGAGAGGACACAGAGCGCTGGGCTC 626
 QY 447 ATGGCAGGATTTAGTGTGATTTGGCACAATTATATCCCA-----TTTATCCC 500
 DB 627 TTGAGCGCGCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 686
 QY 501 CAATGTATCAGTACGATTAAGATTAATTTGCAACAGTGTGACAGATTGACATTTGTA 560
 DB 687 CGAGATCTCGTCCAGGCGCGGACAGCTTCAACATCTGGCCAGGCGGCGCACGCGC 746
 QY 561 TGCCCAAAAAAAGATAGATCTGGGTTGATGTTGCACTGCAATTTATGCTGATGT 620
 DB 747 TGGCAGGATTAAGTGGGTCGCGCTTCAATGTTGACAGTTCATTCAGCGCTTCGTCT 806
 QY 621 ATATGAAGATTTAGCAGCTTTGATTAATGACGTGTTCA-----GGTCTGAAGAGTG 676
 DB 807 CTAAAGACGCTTCCCTCGATTCGAACTAGTGAGTGGTGAACCTGTTCTCCAGGCTT 866
 QY 677 ATCTGGAAGTTCCCAAGATGGAATAAATGATTTGAAGTAACTAGTGG-----A 728
 DB 867 CGAAGACGCTGTTTCGGAGTCGTAAGAGAGCGCGACCCGACATCCGAGGACACGGA 926
 QY 729 ATTCAAAATGAAGATGATACATTAACATACGAAATCAAGTTATTAATGGGTGACGTCAA 788

DB 927 ATGCTAGACTTCGGCATGAGAGCTCCCGCGAATGCAATATGCTCTGTGACGTGCA 986
 QY 789 GGGTGGCTCAGAAACACCCCAATTTGATACAGAGTACTCCAAATGAGAAAAAGAGCC 848
 DB 987 ATGTGTTCCAGACTTCATCATGTGAGAAAGTTTGGAGTGGCGGAAACAGATCA 1046
 QY 849 AGAAGAAAGCTCTGTTGTGTATGACAGCTTAATAGTCCCAATTTACGTTATGAAGA 908
 DB 1047 GAAGGAAGCGGATATGCTCTGGGCGCTGTGCAATCGAACAAGAGAGCTTGGCTGGA 1106
 QY 909 ATTGAGGAAATGCGTGAATAAATACGCTGACAGCCGAGACCTTATTAAGAGTTAGA 968
 DB 1107 ACTCAGACGCTTGGCACAGAGCCCGAGCAATCTCACTGATCTTAATAAATGTTCCG 1166
 QY 969 TCATTCTGTGAGGCTTTGACTGTTGGATTAAGAACATCAGAAAGAGGTTACAGCAAT 1028
 DB 1167 CACCTATATTCAGGCTCG-----CGTACACATCCGTTCCAT 1205
 QY 1029 AACCAAAAATCAGAGGTTCCAAATTGAACCTGATGTCGAAACCAATTGTTGACGTTG 1088
 DB 1206 GACTCAAAAGTCGATGCTCCAAATCGAGCGCGCTCAAAACGAGCTACTTGAAGCTCT 1265
 QY 1089 TCAAGAGATTTCTGTTGT 1148
 DB 1266 GTCCGAGCTGAGGAGGTGATCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1325
 QY 1149 AGCTGATTAAGTGTGGAATC 1171
 DB 1326 TGTGCTCTCATCAAGACATC 1348
 RESULT 37
 ABT18048
 ID ABT18048 standard; DNA; 3508 BP.
 AC ABT18048;
 XX 16-APR-2003 (first entry)
 XX
 DE *Aspergillus fumigatus* essential gene #406.
 XX
 XX Fungicide; cyrostatic; essential gene; *Aspergillus fumigatus*; infection;
 KW cancer; contamination; biofilm; antibody; immune response; ds.
 XX
 OS *Aspergillus fumigatus*.
 XX
 PN WO200286090-A2.
 XX
 PD 31-OCT-2002.
 XX
 PF 23-APR-2002; 2002WO-US013142.
 XX
 PR 23-APR-2001; 2001US-0285697P.
 PR 27-APR-2001; 2001US-0287066P.
 PR 05-JUN-2001; 2001US-0295890P.
 PR 09-JUL-2001; 2001US-0303899P.
 PR 31-AUG-2001; 2001US-0316362P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 DR WPI; 2003-093124/08.
 XX New purified or isolated nucleic acids of essential genes of *Aspergillus*
 PT fumigatus, useful for treating or preventing infections by *A. fumigatus*,
 PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX Disclosure; Page: 175pp; English.
 PS
 XX The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of

interaction. The polypeptides may be used to raise antibodies or to elicit immune response, as a reagent in assays designed to quantitatively determine levels of the protein in biological fluids, as a marker for host tissues in which pathogenic organism invade or reside, and to isolate correlative receptors or ligands in the case of virulence factors. This polynucleotide sequence represents one of the essential genes of *Aspergillus fumigatus* of the invention

XX Sequence 3509 BP, 857 A, 917 C, 883 G, 852 T, 0 U, 0 Other;

Query Match 4.8%; Score 62.6; DB 8; Length 3509;
Best Local Similarity 46.7%; Pred. No. 4,9e-06;
Matches 375; Conservative 0; Mismatches 389; Indels 39; Gaps 4;

```

QY 387 AAAAAATTCTTTACCATTCCTGTCCTTACCGAAGTGGAAAAAGCCGATTAGTTTC 446
DB 1567 AAGCAGATTCTGTGACCTTGTGTTCCTTCAGAGGACCAAGACAGGCTGGGCTC 1626
QY 447 ATCGGAGAGATTAGTGTGAGTTGGCCACAGATTATATCCCA-----TTTATCCC 500
DB 1627 TTACGCGGCTTGTGACTGCTCCCTGTGTCTCTCTGTCATCCACCGTATGCAACC 1686
QY 501 CATGTTATCAGTACGATTAAGATTAATTTTGGACAACGTTGACAGATTGACATTGTTA 560
DB 1687 CGACGATCTCGGTCCAGCGCCGACAAAGCTTCAATCTGGCCAGGCGCCACTGCGC 1746
QY 561 TGGCCAAAAAAGATAGATCTGGGTTTATGTTGAACTGCAATTATAGTCTGATTTC 620
DB 1747 TGGCAGGGTAAAGTCGGGTCCGGCTTCATGTGAGCTGCATTTAAGGCTCTGTC 1806
QY 621 ATATAGAGATTTCACGAGCTTGTGAATAGAGCTGTTTCA-----GTTCTAGAAAGTG 676
DB 1807 CTACAGACCTTCTCCCTCCATTCGAAATCAGTGGGTGACGCTGTTCTCCAGGCTT 1866
QY 677 ATCTGAGAAGTCTCCACAGAGTTGAAAAAATTGATTAAGTAATCTGG-----A 728
DB 1867 CGAAGAGCGTGTGTCGAGTGTAGAGACGCGCACCCCGACATCCGTGGGACACCGA 1926
QY 729 ATTCAAAACATGAAGATGATACATTACCGAATACAGATTATTAATGGGTAGCTGCA 788
DB 1927 ATGTCTAGACTTCGGATGAAGCTCCCGCGAATGCAAAATGCTCTGTGACGTCGA 1986
QY 789 GGGTGGCTCAGAAACACCCAAATTTGTATCACAGTACTCCATGGAAGAAAAAGGAAAGCC 848
DB 1987 ATGTGTTGGCAGATCTCCATCCATGATGAGAAAGTTTGGAGTGGGAGAAACAGATCA 2046
QY 849 AGAAGAAAGCTCTGTGTGTATGACCACTTAATAGTCCCAATTTCAGTTTATGAAGA 908
DB 2047 GAAGGAAGCCGATATGCTCTGGGGCGCTCTGCAATCGAACACAGAGACTTGGCCTGA 2106
QY 909 ATTGAGGGAATATCGTGAATAATATACGACTCAGACCCAGACTTATATTAAGATTAGA 968
DB 2107 ACTCAGACGTTTGCACAGAGCCCGAGCAACTCTCTCAGTGAATTTGAAAATGTCG 2166
QY 969 TCATTCTGTGACCTTTGACTGTGTGATTAAGAACATCAGAAAAGGTTTACAGCAAT 1028
DB 2167 CACTATATTCAGGGCTCG-----GTAACCAATCCGTTCCAT 2205
QY 1029 AACAACAAAATCAGAGTTTCAATTGAACCTGATGTCAAACCAAGTTTGGACCGTTG 1088
DB 2206 GACTCAAAAGTCGATGTCCTCAATCGACGCGCGCTCCAAACCGAGTTACTTTCAGCTCT 2265
QY 1089 TCAAGAGATTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1148
DB 2266 GTCCGAGCTGAGAGGTCTCATCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2325
QY 1149 AGCTGATTAAGTGTGAAAAATC 1171
DB 2326 TGTGCTCTCATCAAGACAATC 2348

```

RESULT 39
ABZ51452

```

ID ABZ51452 standard; cDNA; 869 BP.
XX
AC ABZ51452;
XX
DT 28-MAR-2003 (first entry)
XX
DB Aspergillus oryzae polynucleotide SEQ ID NO 565.
XX
DB Aspergillus oryzae; fermentation; fungus; industrial; EST;
XX expressed sequence tag; gene; ss.
XX
KM Aspergillus oryzae.
XX
OS Aspergillus oryzae.
XX
PN MO200279476-A1.
XX
PD 10-OCT-2002.
XX
PF 22-MAR-2002; 2002MO-IB000890.
XX
PR 30-MAR-2001; 2001JP-00098371.
XX
PA (NAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX
PA (NARE-) NAT RES INST BREWING.
XX
PA (NORQ) NAT FOOD RES INST MIN AGRIC.
XX
PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;
XX Takemuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;
XX WPI, 2003-046817/04.
XX
PT Detection of expression of specific Aspergillus genes for monitoring the
PR fermentation and growth conditions of the fungus, using DNA probes.
XX
PS Claim 1; SEQ ID NO 565; 48bp + Sequence Listing; Japanese.
XX
XX The invention relates to a polynucleotide having any of 6006 specific
CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under
CC specific culture conditions including one or more of eutrophic,
CC oligotrophic, solid, early germination, alkaline, high temperature, low
CC temperature or maltose culture or polynucleotides stringently hybridizing
CC to these sequences. The polynucleotides are useful for monitoring the
CC progress of fermentation and the growth conditions of a fungus,
CC especially of Aspergillus oryzae which is widely used in industrial
CC fermentation. Also monitoring for fungal contamination. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 869 BP; 193 A; 226 C; 239 G; 211 T; 0 U; 0 Other;
XX
Query Match 4.1%; Score 53.2; DB 8; Length 869;
Best Local Similarity 53.5%; Pred. No. 0.00092;
Matches 137; Conservative 0; Mismatches 113; Indels 6; Gaps 1;
QY 390 AACATTCTTTACCATTCCTGTCCTTACCGAAGTGGAAAAAGCCGATTAGTTTATC 449
DB 339 AAGATTCGGAACCTTCGGTGTCTCTTTCACAGAGCCCAAGACAGGCTAGTGTCTC 398
QY 450 GCGAGATTAGTGTGAGTTGTTGCCAAGTTTATATCCATTATC-----CCCA 503
DB 399 TGGGCTCTAGTACTGCTCCCTAGTATCATCCCTGTTATTCACCGTACCTGACGCTGA 458
QY 504 TGTATCAGTACGATTAAGATTAATTTTGGACAACGTTGACAGATTGACATTGTTATGC 563
DB 459 CGACCTTGAGGCTTCTCGTGAACAAGCTTCATTAATCTTGACAGGCTGCCCACGTGCTGC 518
QY 564 CCAAAAAAAGATAGATCTGGGTTTATGTTTGCACACTGCAATTATAGTCTGATTATATA 623
DB 519 TCAAGGTAAAGTGGAGTCCGGGTTTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 578
QY 624 TGAAGATTTCAGCCA 639
DB 579 TGGCGAATTCCTCCCA 594

```

```

RESULT 40
ADA71938/C
ID ADA71938 standard; DNA; 2000 BP.
XX
XX ADA71938;
AC
XX 20-NOV-2003 (first entry)
DT
XX
XX
DE Rice gene, SEQ ID 5263.
KM Plant; bacterial infection; fungal infection; viral infection; rice;
KM gene; ds.
XX
XX Oryza sativa.
OS
XX WO2003000898-A1.
PN
XX 03-JAN-2003.
PD
XX 22-JUN-2001; 2001WO-1B001105.
PF
XX 22-JUN-2001; 2001WO-1B001105.
PR
XX 22-JUN-2001; 2001WO-1B001105.
XX
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
PA
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitcham S, Xie Z, Zhu T, Zou G;
XX WPI; 2003-175290/17.
DR
XX
XX Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
XX Claim 27; SEQ ID NO 5263; 899bp; English.
XX
XX The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
XX Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;
SQ
Query Match 3.7%; Score 48.2; DB 8; Length 2000;
Best Local Similarity 6.8%; Pred. No. 0.028;
Matches 52; Conservative 372; Mismatches 340; Indels 3; Gaps 2;
Oy 18 TGCACCTGGAAAGACATTTCTTGCGTGGATATTTGTTCTTGAGCCAATTATGATGC 77
Db 1018 TGMHTTTRSMWYTAAMKKTKTKTMTAVSSTWKMAYRAVWSRSRKTWCTGGKRM 959
Oy 78 TTATGTGACAGATGTGTCATCAGATGCATGATGATTAACCAAGAACACAGTTT 137
Db 958 ATYCGTKMAAAGRWBRMAWMCWCCMKMKWKTSCMMWXYWRTWSCWYTMWNGAMRYAY 899
Oy 138 GAAAGATCTGAATCAAAATTTCTTACCCCAATTTGCAAGCGAAGTGAATATCA 197
Db 898 AWRRRRWYTKWBRWYTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMT 839
Oy 198 CATATCATCAATPACAGAGAGCCAGAGATTCAGTCAAGCATTAATCCATTTTGA 257
Db 838 CAKCYTKMAATWMTTWAACAPATSRWRBAMGMRKRYKMKRDAYWRRWRRCWKAQWAW 779
Oy 258 GGCAACTATATTCATCGTTTGTAGCTTATATTCACCGACCGAAGCATTTGATCTTGAAT 317

```

```

Db 778 MRS-RYRMKMKYATRYWKMAMTWMSWRWKSYRMWSGWRWMSAMRYCSRKCKAK 720
Oy 318 CATCATTTTACTGACCCCTGATATCATCACAAGAAAGATGAAACAACATCTCTC 377
Db 719 TKYASARWTKAASRYRRRRWYKWKGTWYRYWRSRBTTRAMSRRKKAASMS 660
Oy 378 GAAT--GGAGAAAAAATCTTTTCAATTCCTGTCGTCATTAACGAAAGTGAAGAAC 435
Db 659 CWWYRGAASWYSKYSCAKCKKTRWYVTSYSMTGMYGMSYSYKSMWTSKMSYGM 600
Oy 436 GGATTAAGTTCATCGCAGAGATTAGTCAGTTTGCCACAAGTTTATATCCCATTTT 495
Db 599 TCTMYTSMKGSYRBRKMGWMSGMSRMYRMWKRKRYRMYKWKCTRRCMCYRWGYT 540
Oy 496 ATCCCAATGTATCAGTACGAATAAGATATTTTGCACACGTCGACAGATTGCACAT 555
Db 539 MTTTSSRMWMTGRYKARTSKRRYMYKYRKICYIYGYMKKCSYMMRYGICAKCKK 480
Oy 556 TGTATGCCCCAAAAAAGATAGATCTGGTTTGATGTGCAACTGCAATTTATGCTCTG 615
Db 479 CYAMCKAAVSGMMWYVYRKYSKMRBMSWKYMSWYKCRSMKYGAKGCGCKMTTYS 420
Oy 616 ATGTATATAGAAGATTTTCAAGCAGCTTTGATTAATGACGTTTCAGGTTCTAGAAAT 675
Db 419 YGYMKWYTYMGSYKYSRCYKWRMYMYKGMWYMYSAVSSMTWYVYAKWYKRR 360
Oy 676 GATCCTGGAATTCGCCACAGAGTTGAAAAAATGATGAAGTAAGTGGGAATTCAA 735
Db 359 GTMSWYGSYKRYKCTWMCYMKCMRCYRWRKWKTKYSKRCYCRVATCYWCCYRKR 300
Oy 736 CATGAAGATGTACATTAACATACGAAATCAAGTATTTAATGGTGA 782
Db 299 GYSSRSMMRTAGKMKMSWSRWCYSYVYKMKMKSYMSYSGW 253

```

Search completed: January 26, 2005, 04:43:00
Job time : 496 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 26, 2005, 06:40:30 ; Search time 3871 Seconds
(without alignments)
15869.109 Million cell updates/sec

Title: US-10-069-062-6
Perfect score: 1299
Sequence: 1 atgtcaaaagcatttctgtc.....aagactatatagttctataa 1299

Scoring table: OLIGO_NUC
Gapop_60.0, Gapext 60.0

Searched: 4526729 seqs, 23644849745 residues

Word size: 0

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database:

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_on:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_rsb:*
12: gb_by:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1299	100.0	1299	6 AX087879	Sequence
2	1299	100.0	1763	6 AX087878	Sequence
3	891	68.6	1299	6 AX240804	Sequence
4	891	68.6	1299	6 AX488723	Sequence
5	421	32.4	547	6 AX087874	Sequence
6	334	25.7	577	6 AX087876	Sequence
7	65	5.0	65	6 AX485723	Sequence
8	65	5.0	25	6 AX241110	Sequence
9	25	1.9	25	6 AX087877	Sequence
10	24	1.8	36	6 AX087884	Sequence
11	23	1.8	110000	10 AE014174_2	Continuation (3 of
12	23	1.8	201532	2 AC115701	Mus muscu
13	22	1.7	22	6 AX241232	Sequence
14	22	1.7	22	6 AX487723	Sequence
15	22	1.7	58879	2 AC090350	Sequence
16	22	1.7	153598	2 AC149142	Xenopus t
17	22	1.7	171133	9 AC103846	Homo sapi
18	22	1.7	185418	2 AC012311	Homo sapi
19	22	1.7	188270	9 AC104393	Homo sapi

20	22	1.7	192347	9 AC007614	Homo sapi
21	21	1.6	21	6 AX087875	Sequence
22	21	1.6	313	6 C061526	Sequence
23	21	1.6	479	6 AX396511	Sequence
24	21	1.6	530	6 AX397039	Sequence
25	21	1.6	771	6 BD020934	Sequence
26	21	1.6	771	6 BD100872	Sequence
27	21	1.6	1444	6 CQ117294	Sequence
28	21	1.6	1596	4 AF068483	Oryctolag
29	21	1.6	1685	4 HMMHTR3A	Human chape
30	21	1.6	2010	9 HMMTCR20	Human chape
31	21	1.6	2562	9 AF385084	Homo sapi
32	21	1.6	2647	9 AB063318	Homo sapi
33	21	1.6	2837	6 CQ492936	Sequence
34	21	1.6	23130	2 AC020045	Drosophila
35	21	1.6	38875	9 AC092101	Homo sapi
36	21	1.6	115815	8 AC114896	Homo sapi
37	21	1.6	128524	2 AC012253	Homo sapi
38	21	1.6	128685	10 AL772361	Mouse DNA
39	21	1.6	134105	5 AC096845	Mouse DNA
40	21	1.6	145042	10 AL732425	Mouse DNA
41	21	1.6	152927	9 AC004895	Homo sapi
42	21	1.6	153174	2 AC135904	Rattus no
43	21	1.6	154791	2 AC069393	Homo sapi
44	21	1.6	164669	2 AC140966	Sus scrofa
45	21	1.6	174394	2 AC112725	Drosophila
46	21	1.6	183338	2 AC083775	Homo sapi
47	21	1.6	187772	3 AC104628	Drosophila
48	21	1.6	188914	9 AC093744	Homo sapi
49	21	1.6	189976	3 AC023710	Drosophila
50	21	1.6	191291	2 AC148313	Pan trogl
51	21	1.6	197639	9 AC009554	Homo sapi
52	21	1.6	198684	2 AC148309	Pan trogl
53	21	1.6	201932	5 BX897729	Zebrafish
54	21	1.6	202816	2 CR407544	Danio rer
55	21	1.6	215435	10 AL662838	Mouse DNA
56	21	1.6	215972	2 AC146017	Pan trogl
57	21	1.6	217022	2 AC146146	Pan trogl
58	21	1.6	232385	2 AC133770	Rattus no
59	21	1.6	238009	2 AC103650	Mus muscu
60	21	1.6	242109	2 AC128430	Rattus no
61	21	1.6	242467	2 BX927324	Danio rer
62	21	1.6	305376	3 AE003437	Drosophila
63	20	1.5	33	6 AX087883	Sequence
64	20	1.5	630	6 AX505906	Sequence
65	20	1.5	630	6 BX006211	Arabidops
66	20	1.5	919	8 AK117303	Arabidops
67	20	1.5	1476	8 AY113176	Arabidops
68	20	1.5	1535	3 SAR31MSA	Sarcocystis
69	20	1.5	1978	8 AY054231	Arabidops
70	20	1.5	2490	8 BT003013	Arabidops
71	20	1.5	2856	8 AY125524	Arabidops
72	20	1.5	3527	5 XU26256	Xenopus lae
73	20	1.5	3585	5 BC059339	Xenopus lae
74	20	1.5	14745	9 PTEN3	Human DNA
75	20	1.5	39487	9 AC148869	Pan trogl
76	20	1.5	52717	8 AB019227	Arabidops
77	20	1.5	62580	2 AL355143	Human DNA
78	20	1.5	63710	2 AC084183	Homo sapi
79	20	1.5	74560	2 HS433B8	Human DNA
80	20	1.5	89582	9 AC022429	Homo sapi
81	20	1.5	100000	9 AP000093	Homo sapi
82	20	1.5	100000	9 AP000197	Homo sapi
83	20	1.5	107515	2 AC147403	Dasyatis n
84	20	1.5	107603	2 AC016661	Arabidops
85	20	1.5	110000	2 AC123648_2	Continuation (3 of
86	20	1.5	110000	2 AC123879_1	Continuation (2 of
87	20	1.5	110000	2 AC140127_2	Continuation (3 of
88	20	1.5	113126	8 AC007576	Arabidops
89	20	1.5	115348	8 AL357042	Human DNA
90	20	1.5	116166	9 AC024109	Homo sapi
91	20	1.5	118439	9 AL135905	Human DNA
92	20	1.5	120045	9 AL353052	Human DNA

AC007614	Homo sapi
AX087875	Sequence
C061526	Sequence
AX396511	Sequence
AX397039	Sequence
BD020934	Novel gen
BD100872	Novel gen
CQ117294	Sequence
AF068483	Oryctolag
M94083	Human chape
L27706	Human chape
AF385084	Homo sapi
AB063318	Homo sapi
CQ492936	Sequence
AC020045	Drosophila
AC092101	Homo sapi
AC114896	Oryza sat
AC012253	Homo sapi
AL772361	Mouse DNA
AC096845	Mouse DNA
AL732425	Mouse DNA
AC004895	Homo sapi
AC135904	Rattus no
AC069393	Homo sapi
AC140966	Sus scrofa
AC112725	Drosophila
AC083775	Homo sapi
AC104628	Drosophila
AC093744	Homo sapi
AC023710	Drosophila
AC148313	Pan trogl
AC009554	Homo sapi
AC148309	Pan trogl
BX897729	Zebrafish
CR407544	Danio rer
AL662838	Mouse DNA
AC146017	Pan trogl
AC146146	Pan trogl
AC133770	Rattus no
AC103650	Mus muscu
AC128430	Rattus no
BX927324	Danio rer
AE003437	Drosophila
AX087883	Sequence
AX505906	Sequence
BT006211	Arabidops
AK117303	Arabidops
AY113176	Arabidops
M76496	Sarcocystis
AY054231	Arabidops
BT003013	Arabidops
AY125524	Arabidops
U26256	Xenopus lae
BC059339	Xenopus lae
PTEN3	Human DNA
AC148869	Pan trogl
AB019227	Arabidops
AL355143	Human DNA
AC084183	Homo sapi
HS433B8	Human DNA
AC022429	Homo sapi
AP000093	Homo sapi
AP000197	Homo sapi
AC147403	Dasyatis n
AC016661	Arabidops
AC123648_2	Continuation (3 of
AC123879_1	Continuation (2 of
AC140127_2	Continuation (3 of
AC007576	Arabidops
AL357042	Human DNA
AC024109	Homo sapi
AL135905	Human DNA
AL353052	Human DNA

93	20	1.5 120112	9	AC124850	166	20	1.5 211843	2	AC145094	AC145094 Homo sapi
94	20	1.5 120527	8	AC141114	167	20	1.5 211884	2	AC139790	AC139790 Homo sapi
95	20	1.5 125529	2	AC121242	168	20	1.5 213388	2	AC138920	AC138920 Homo sapi
96	20	1.5 128039	2	AC010421	169	20	1.5 213387	2	AC141595	AC141595 Homo sapi
97	20	1.5 130630	2	AC010637	170	20	1.5 215130	5	AL954320	AL954320 Zebrafish
98	20	1.5 132403	2	AC148792	171	20	1.5 215682	10	AC107369	AC107369 Mus muscu
99	20	1.5 133561	9	AL951044	172	20	1.5 218336	9	AF067844	AF067844 Homo sapi
100	20	1.5 134091	2	AC023349	173	20	1.5 222040	2	AC098532	AC098532 Rattus no
101	20	1.5 134722	8	AC137923	174	20	1.5 226382	2	AC097295	AC097295 Rattus no
102	20	1.5 139121	2	AL135916	175	20	1.5 227049	2	AC079653	AC079653 Mus muscu
103	20	1.5 140871	2	AC016867	176	20	1.5 227487	2	AC111354	AC111354 Rattus no
104	20	1.5 146032	2	AC145253	177	20	1.5 228353	2	AL445283	AL445283 Homo sapi
105	20	1.5 152445	9	AC138410	178	20	1.5 229447	2	AC123228	AC123228 Rattus no
106	20	1.5 153265	9	AP000237	179	20	1.5 235468	10	AC094489	AC094489 Rattus no
107	20	1.5 153667	2	AC087145	180	20	1.5 239901	2	AC125552	AC125552 Rattus no
108	20	1.5 154311	9	AF107258	181	20	1.5 241629	2	AC136278	AC136278 Rattus no
109	20	1.5 154891	9	AC104131	182	20	1.5 242931	2	AC107748	AC107748 Rattus no
110	20	1.5 160582	9	AC105429	183	20	1.5 243366	2	AC150537	AC150537 Rattus no
111	20	1.5 160653	2	AC026306	184	20	1.5 243920	2	AC117886	AC117886 Rattus no
112	20	1.5 161460	9	AC093243	185	20	1.5 249520	2	AC096833	AC096833 Rattus no
113	20	1.5 162964	9	AL157879	186	20	1.5 252771	2	AC129788	AC129788 Rattus no
114	20	1.5 163031	9	AC011939	187	20	1.5 252311	2	AC135471	AC135471 Rattus no
115	20	1.5 163311	2	AC022016	188	20	1.5 262324	2	AC096148	AC096148 Rattus no
116	20	1.5 163311	2	AC147402	189	20	1.5 302422	1	AE017011	AE017011 Bacillus
117	20	1.5 164308	2	AC140143	190	20	1.5 302422	1	AC145103	AC145103 Homo sapi
118	20	1.5 164598	2	AC138938	191	20	1.5 340000	9	AP001706	AP001706 Homo sapi
119	20	1.5 165477	2	AC138944	192	19	1.5 340000	9	HS75E1R	HS75E1R
120	20	1.5 167798	2	AC138044	193	19	1.5 247	11	BV088038	BV088038 sc1352_p3
121	20	1.5 168173	2	AC138947	194	19	1.5 255	11	BV088042	BV088042 sc1352_p3
122	20	1.5 168547	2	AC138947	195	19	1.5 339	11	BV116618	BV116618 PZA00783
123	20	1.5 169082	2	AC138947	196	19	1.5 406	11	BV116623	BV116623 PZA00783
124	20	1.5 170044	2	AC137696	197	19	1.5 414	11	BV116625	BV116625 PZA00783
125	20	1.5 170264	2	AC137640	198	19	1.5 452	3	AY363002	AY363002 Colieter
126	20	1.5 170816	2	AC138411	199	19	1.5 623	11	BV146882	BV146882 PZA02654-
127	20	1.5 170928	9	AC148803	200	19	1.5 646	11	BV146887	BV146887 PZA02654-
128	20	1.5 171705	2	AC139458	201	19	1.5 647	11	BV146880	BV146880 PZA02654-
129	20	1.5 175447	9	AC091607	202	19	1.5 648	11	BV146879	BV146879 PZA02654-
130	20	1.5 175942	2	AC140014	203	19	1.5 648	11	BV146888	BV146888 PZA02654-
131	20	1.5 176010	2	AC062925	204	19	1.5 656	11	BV146884	BV146884 PZA02654-
132	20	1.5 176189	2	AC062925	205	19	1.5 657	11	BV146881	BV146881 PZA02654-
133	20	1.5 177462	2	AC133283	206	19	1.5 657	11	BV146883	BV146883 PZA02654-
134	20	1.5 178464	2	AC140012	207	19	1.5 659	11	BV146886	BV146886 PZA02654-
135	20	1.5 178805	2	AC138858	208	19	1.5 661	11	BV146878	BV146878 PZA02654-
136	20	1.5 179932	2	AC079563	209	19	1.5 669	11	BV146885	BV146885 PZA02654-
137	20	1.5 182173	2	AC079563	210	19	1.5 1107	8	AK067291	AK067291 Oryza sat
138	20	1.5 182991	2	AC121349	211	19	1.5 1242	8	MCU80070	MCU80070 Mesembryant
139	20	1.5 183155	9	BS000175	212	19	1.5 1282	6	CQ492059	CQ492059 Sequence
140	20	1.5 183694	9	AC005412	213	19	1.5 1282	6	CQ497971	CQ497971 Sequence
141	20	1.5 183815	9	AL451084	214	19	1.5 1471	3	AY363004	AY363004 Colieter
142	20	1.5 184624	2	AC138868	215	19	1.5 2069	5	AE031243	AE031243 Lotus jap
143	20	1.5 185096	9	AL355140	216	19	1.5 2085	5	CH389521	CH389521 Gallus ga
144	20	1.5 185220	2	AC140173	217	19	1.5 2649	8	AK100281	AK100281 Oryzo sat
145	20	1.5 186595	10	AC140053	218	19	1.5 3471	8	SCYGL049C	SCYGL049C S. cerevisia
146	20	1.5 187740	9	AC138409	219	19	1.5 3900	8	YECTIF4632	YECTIF4632 Saccharomyc
147	20	1.5 187939	5	AL928672	220	19	1.5 4265	1	AF458777	AF458777 Lactococc
148	20	1.5 188179	2	AC139286	221	19	1.5 4532	14	AY029520	AY029520 F.lji dise
149	20	1.5 188352	2	AC139496	222	19	1.5 5249	1	AE006527	AE006527 Lactococc
150	20	1.5 188642	9	HS55C20	223	19	1.5 5749	10	PMU0288129	PMU0288129 Mus muscu
151	20	1.5 190194	9	AC068025	224	19	1.5 6490	8	SPC1215	SPC1215 S. pombe c
152	20	1.5 191957	9	AC015845	225	19	1.5 6733	6	CQ612740	CQ612740 Sequence
153	20	1.5 193309	2	AC138958	226	19	1.5 10769	6	CQ830738	CQ830738 Sequence
154	20	1.5 193623	2	AC138820	227	19	1.5 11137	1	AE003868	AE003868 Sequence
155	20	1.5 196264	2	AC139780	228	19	1.5 12181	6	CQ612728	CQ612728 Sequence
156	20	1.5 199001	2	AL157880	229	19	1.5 13350	3	CHRG46N07	CHRG46N07 Caenorhab
157	20	1.5 201462	2	AC139283	230	19	1.5 14674	9	AL513490	AL513490 Human DNA
158	20	1.5 201536	9	AC063965	231	19	1.5 16336	9	AC034208	AC034208 Homo sapi
159	20	1.5 202679	2	AC141260	232	19	1.5 30370	8	AP006078	AP006078 Homo sapi
160	20	1.5 203221	10	AC102689	233	19	1.5 60464	2	AL959673	AL959673 Homo sapi
161	20	1.5 204533	2	AC116994	234	19	1.5 62871	2	AC124264	AC124264 Homo sapi
162	20	1.5 204617	2	EX324160	235	19	1.5 62871	2	AC124264	AC124264 Homo sapi
163	20	1.5 207430	10	AL772236	236	19	1.5 64036	2	AC040921	AC040921 Homo sapi
164	20	1.5 207459	9	AC139792	237	19	1.5 70574	9	AC017050	AC017050 Homo sapi
165	20	1.5 210040	2	AC107603	238	19	1.5 70574	9	AC017050	AC017050 Homo sapi

AP001971	Homo sapi
AC037197	Oryza sat
AC0012110	Homo sapi
AP004337	Oryza sat
AC134053	Oryza sat
AC134053	Oryza sat
AL005927	Mus muscu
AC087130	Mus muscu
BC294194	Mus muscu
AL105840	Homo sapi
AL152492	Homo sapi
AP005126	Oryza sat
AC088994	Felis cat
AC079357	Oryza sat
AC065222	Homo sapi
AC001420	Homo sapi
AL391495	Homo sapi
AL391495	Homo sapi
AL312075	Felis cat
AP003507	Oryza sat
AC104272	Oryza sat
AL162725	Homo sapi
AC025188	Homo sapi
CC006062	Homo sapi
AC164653	Salimira s
AP002839	Oryza sat
AP003032	Oryza sat
AP004854	Oryza sat
AL1513487	Oryza sat
AC132002	Human DNA
AC001630	Mus muscu
AP005129	Oryza sat
AC001510	Homo sapi

LOCUS	1299 bp	DNA	PAT 17-MAR-2001
AX087879	Sequence 6 from Patent WO0114533.		linear
AX087879	ACCESSION		
AX087879	DEFINITION		
AX087879	SEQUENCE		

```
/organism="Candida albicans"  
/mol_type="unassigned DNA"  
/db_xref="taxon:5476"
```

```

Mismatches 0; Indels 0; Gaps 0;

```

AGTTTGAAGAATCTAGAAATCAAATTTCTTACCCCAATTGGCAAC 180
GATGCTTATGTGACACGATTTGTCAATCGAATGCATGCAGTTATAACA 120

Db	121	CCAAAAGAACCAAGTTTGAAGAACTGAAATCAAAAATTTCTTCAACCCCAATTTGGCAAC	180
Qy	181	GGAGATGGGAATATCA CATATCATCAAAATACAGAGAGCCAGAGAAAGTTCACTCAACG	240
Db	181	GGAAATGGGAATATCA CATATCATCAAAATACAGAGAGCCAGAGAAAGTTCACTCAACG	240
Qy	241	ATAAATCATTTTAAAGGGAACATATTCATAGCTTTTAACTTAATATTCACCGACGAA	300
Db	241	ATAAATCATTTTAAAGGGAACATATTCATAGCTTTTAACTTAATATTCACCGACGAA	300
Qy	301	GCAATTTGATCTTGAATATCATCATTTTACTGACCCCTGGATATCATTCACAGAAATACT	360
Db	301	GCAATTTGATCTTGAATATCATCATTTTACTGACCCCTGGATATCATTCACAGAAATACT	360
Qy	361	GAAGCCAAAGCATCTTCGAAATGAGAAAAACATTTCTTTACATTCCTGTCATTAAC	420
Db	361	GAAGCCAAAGCATCTTCGAAATGAGAAAAACATTTCTTTACATTCCTGTCATTAAC	420
Qy	421	GAAGTGAAGAAAGACCGAATTAAGTTGATCGGCGAGATTAAGTGCATCTGTCACAAGT	480
Db	421	GAAGTGAAGAAAGACCGAATTAAGTTGATCGGCGAGATTAAGTGCATCTGTCACAAGT	480
Qy	481	TTATTTATCCCATTTTATCCCAATGTTATCAGTACGAATTAAGATATTTTGGACAAGTT	540
Db	481	TTATTTATCCCATTTTATCCCAATGTTATCAGTACGAATTAAGATATTTTGGACAAGTT	540
Qy	541	GCACAGATTTGCA CATTTGTTATGCCAAAAAAGATAGATCTGGGTTTGAATTTGCACT	600
Db	541	GCACAGATTTGCA CATTTGTTATGCCAAAAAAGATAGATCTGGGTTTGAATTTGCACT	600
Qy	601	GCAATTTATGCTGTGATTTATATAGAAATTTCAAGCAGCTTTGATTAATACGTTGTT	660
Db	601	GCAATTTATGCTGTGATTTATATAGAAATTTCAAGCAGCTTTGATTAATACGTTGTT	660
Qy	661	CAGGTTCTAGAAAGATCTGAGAAAGTTCCCAAGATTTGAAAAAATGATGAAGT	720
Db	661	CAGGTTCTAGAAAGATCTGAGAAAGTTCCCAAGATTTGAAAAAATGATGAAGT	720
Qy	721	AACTGGGAATTCAAA CATGAAGAAGATGTACATTA CCAATAGGAATCAATTTATATGGGT	780
Db	721	AACTGGGAATTCAAA CATGAAGAAGATGTACATTA CCAATAGGAATCAATTTATATGGGT	780
Qy	781	GACGTCAGAGGGTGTCA GAAACACCCAAATTTGATACAGATCTTCAAATGAAAAAG	840
Db	781	GACGTCAGAGGGTGTCA GAAACACCCAAATTTGATACAGATCTTCAAATGAAAAAG	840
Qy	841	GAAGAGCCAGAAAGAGCTGTGTGTGTGATGACAGAGTTAATATTCACATTTCACTTT	900
Db	841	GAAGAGCCAGAAAGAGCTGTGTGTGTGATGACAGAGTTAATATTCACATTTCACTTT	900
Qy	901	ATGAAGGATTTGAGGGAATTCGTGA AAAATTCAGACTCAGACCCAGAGACTTATATTA	960
Db	901	ATGAAGGATTTGAGGGAATTCGTGA AAAATTCAGACTCAGACCCAGAGACTTATATTA	960
Qy	961	GAGTTAGATCATCTGTGAGCCCTTTGACGTGCGCATTAAGAACATCAAAAAAGGCTTA	1020
Db	961	GAGTTAGATCATCTGTGAGCCCTTTGACGTGCGCATTAAGAACATCAAAAAAGGCTTA	1020
Qy	1021	CAAGCATTTAACAAAAAATCAGAGGTTCCAAATTTGAACATCTTCAAACCCAGTTGTTG	1080
Db	1021	CAAGCATTTAACAAAAAATCAGAGGTTCCAAATTTGAACATCTTCAAACCCAGTTGTTG	1080
Qy	1081	GACCGTTGTAAGAAAGATCTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATAC	1140
Db	1081	GACCGTTGTAAGAAAGATCTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATAC	1140
Qy	1141	GATGCAATAGCTTATTAAGTTGAGAAAAATCAAGTGGAAAAATTTAAGACAGAAATCTTT	1200
Db	1141	GATGCAATAGCTTATTAAGTTGAGAAAAATCAAGTGGAAAAATTTAAGACAGAAATCTTT	1200
Qy	1201	GAAGATTCAGATTAATTTTCAATATGTTTAACTGGGTGATTTTGGAGAGCAAAACAGAAAGT	1260
Db	1201	GAAGATTCAGATTAATTTTCAATATGTTTAACTGGGTGATTTTGGAGAGCAAAACAGAAAGT	1260

Db	1201	GAAATCCAGATTATTTTCATPAAAGTTTAACTGCGGTGATTGGAAAGACAACAGAGCT	1260
Qy	1261	GTACTTGAAGAAAAACGAAAGACTATATAGGTTATATA	1299
Db	1261	GTACTTGAAGAAAAACGAAAGACTATATAGGTTATATA	1299
RESULT 2			
LOCUS	AX087878	1763 bp	DNA
DEFINITION	Sequence 5 from Patent WO0114533.		linear
ACCESSION	AX087878		
VERSION	AX087878.1	GI:13396871	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	Rosamond, J.D. and Schnell, N.F.		
TITLE	Phosphomevalonate kinase (pmk) gene (erg8) from candida albicans		
JOURNAL	Patent: WO 0114533-A 5 (1-MAR-2001); Astrazeneca AB (SE)		
FEATURES			
source	location/Qualifiers		
	1. 1763		
	/organism="Candida albicans"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:5476"		
ORIGIN			
Query Match	100.0%;	Score 1299;	DB 6; Length 1763;
Best Local Similarity	100.0%;	Pred. No. 0;	
Matches 1299;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Qy	1	ATGTCAAAAGCATTTAGTGCACTGGAAAGAGCATTTCTTGCTGGTGATATTTGGTCTT	60
Db	89	ATGTCAAAAGCATTTAGTGCACTGGAAAGAGCATTTCTTGCTGGTGATATTTGGTCTT	148
Qy	61	GAGCCAAATTTATGATGCTTATGTGCAGCAGCATTTGCATCAGAAATGATGATGATATA	120
Db	149	GAGCCAAATTTATGATGCTTATGTGCAGCAGCATTTGCATCAGAAATGATGATGATATA	208
Qy	121	CCAAAAGAACCGATTGTAAGAAATCTAGATCAAATTTCTTCAACCCCAATTTGCAAC	180
Db	209	CCAAAAGAACCGATTGTAAGAAATCTAGATCAAATTTCTTCAACCCCAATTTGCAAC	268
Qy	181	GGAGAATGGGAATATCATATCATCAATCAAAATCAGAGAACCCAGAGAGTTCAATCAGC	240
Db	269	GGAGAATGGGAATATCATATCATCAAAATCAGAGAACCCAGAGAGTTCAATCAGC	328
Qy	241	ATAAATCCATTTTATAGAGCAACTATATTCATCGTTTATAGCTTATATTCACACCGAA	300
Db	329	ATAAATCCATTTTATAGAGCAACTATATTCATCGTTTATAGCTTATATTCACACCGAA	388
Qy	301	GCATTGTGATCTTGAATATCATATTTACTCAGACCTCGATATCATTCACAAGAGATAT	360
Db	389	GCATTGTGATCTTGAATATCATATTTACTCAGACCTCGATATCATTCACAAGAGATAT	448
Qy	361	GAAACCAAGATCCTCGAATGGAGAAAAACATTTCTTTACATTCCTTCGCGCAATTAC	420
Db	449	GAAACCAAGATCCTCGAATGGAGAAAAACATTTCTTTACATTCCTTCGCGCAATTAC	508
Qy	421	GAAATGAAAAAGACCGATATAGTTTCATCGGACGAGATTAAGTCAAGTTGGTCCACAGT	480
Db	509	GAAATGAAAAAGACCGATATAGTTTCATCGGACGAGATTAAGTCAAGTTGGTCCACAGT	568
Qy	481	TTATTTATCCATTTTATCCCAATGTATCAATACGAATAAAGATATTTTGACACAGTT	540
Db	569	TTATTTATCCATTTTATCCCAATGTATCAATACGAATAAAGATATTTTGACACAGTT	628
Qy	541	GCACAGATTTGACATTTGTTATAGCCAAAAAAGATAGATCTGGGTTTGATGTTGCACT	600
Db	629	GCACAGATTTGACATTTGTTATAGCCAAAAAAGATAGATCTGGGTTTGATGTTGCACT	688

[illegible]

Query Match 68.6%; Score 891; DB 6; Length 1299;
 Best Local Similarity 99.4%; Pred. No. 0;
 Matches 1291; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

QY 1 ATGTCAAAAAGATTAGTACCTGGAAGCATTTCTGTGTGTGATATTGGTCTT 60
DB 1 ATGTCAAAAAGATTAGTACCTGGAAGCATTTCTGTGTGTGATATTGGTCTT 60
QY 61 GAGCCAAATTTATGATGCTTATGTGACAGCATTTGATCAAGATGCAATGCAATTAATACA 120
DB 61 GAGCCAAATTTATGATGCTTATGTGACAGCATTTGATCAAGATGCAATGCAATTAATACA 120
QY 121 CCAAAAGGAACCAATTGAAAAGAAATCTAGATCAAAATTTCTTCAACCCCAATTTGCAAC 180
DB 121 CCAAAAGGAACCAATTGAAAAGAAATCTAGATCAAAATTTCTTCAACCCCAATTTGCAAC 180
QY 181 GGAGATGGGAATGTCAATATCATCAATTCAGAGAGAGCCGAGAGAGTTCACTACCC 240
DB 181 GGAGATGGGAATGTCAATATCATCAATTCAGAGAGAGAGCCGAGAGAGTTCACTACCC 240
QY 241 ATAAATCCATTTTATAGAGCACTATATTCATGCTTTATATTAATCAACCGAGA 300
DB 241 ATAAATCCATTTTATAGAGCACTATATTCATGCTTTATATTAATCAACCGAGA 300
QY 301 GCAATTTGATCTGAATATCATATTTACTGACACCTTGATATCATTCACAAGAGATACT 360
DB 301 GCAATTTGATCTGAATATCATATTTACTGACACCTTGATATCATTCACAAGAGATACT 360
QY 361 GAAACCAAGACATCTCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 361 GAAACCAAGACATCTCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY 421 GAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 421 GAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 481 TTATATCCCATTTTATCCCAATGTTTATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
DB 481 TTATATCCCATTTTATCCCAATGTTTATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
QY 541 GCAAGATTTGACATTTGTTATGCCCCAAAAAGATAGAGATCTGGGTTGATGTTGCAACT 600
DB 541 GCAAGATTTGACATTTGTTATGCCCCAAAAAGATAGAGATCTGGGTTGATGTTGCAACT 600
QY 601 GCAATTTATGCTGTGATGTTATGAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 601 GCAATTTATGCTGTGATGTTATGAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 661 CAGGTTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 661 CAGGTTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 721 AACTGGAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 721 AACTGGAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 781 GACGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 781 GACGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
QY 841 GAAAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 841 GAAAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 901 ATGAAGGAATTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 901 ATGAAGGAATTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
QY 961 GAGTTAGATCATTTCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
DB 961 GAGTTAGATCATTTCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
QY 1021 CAAGCATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
DB 1021 CAAGCATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080

```

```

DB 1021 CAAGCATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
QY 1081 GACCGTTGCAAGAGATTCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
DB 1081 GACCGTTGCAAGAGATTCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
QY 1141 GATGCAATAGCTGTATATGTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1141 GATGCAATAGCTGTATATGTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1201 GAAAATCCAGATTTATTTTCAATATGTTTACTGCGTTGATTTGAGAGAGAGAGAGAGAGAG 1260
DB 1201 GAAAATCCAGATTTATTTTCAATATGTTTACTGCGTTGATTTGAGAGAGAGAGAGAGAGAG 1260
QY 1261 GTACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1299
DB 1261 GTACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1299

```

RESULT 4
 AX488723 1299 bp DNA linear PAT 16-AUG-2002
 LOCUS
 DEFINITION Sequence 6023 from Patent WO02053728.
 ACCESSION AX488723
 VERSION AX488723.1 GI:22322735
 KEYWORDS
 SOURCE
 ORGANISM
 Candida albicans
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE
 1 Roemer, T., Jiang, B., Boone, C., Bussey, H. and Ohlsen, K. L.
 Gene disruption methodologies for drug target discovery
 Patent: WO 02053728-A 6023 11-JUL-2002;
 JOURNAL Elitira Pharmaceuticals, Inc. (US)
 location/Qualifiers
 FEATURES
 source
 1..1299
 /organism="Candida albicans"
 /mol_type="unassigned DNA"
 /db_xref="taxon:5476"

ORIGIN

```

Query Match 68.6%; Score 891; DB 6; Length 1299;  

Best Local Similarity 99.4%; Pred. No. 0;  

Matches 1291; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 ATGTCAAAAAGATTAGTACCTGGAAGCATTTCTGTGTGTGATATTGGTCTT 60
DB 1 ATGTCAAAAAGATTAGTACCTGGAAGCATTTCTGTGTGTGATATTGGTCTT 60
QY 61 GAGCCAAATTTATGATGCTTATGTGACAGCATTTGATCAAGATGCAATGCAATTAATACA 120
DB 61 GAGCCAAATTTATGATGCTTATGTGACAGCATTTGATCAAGATGCAATGCAATTAATACA 120
QY 121 CCAAAAGGAACCAATTGAAAAGAAATCTAGATCAAAATTTCTTCAACCCCAATTTGCAAC 180
DB 121 CCAAAAGGAACCAATTGAAAAGAAATCTAGATCAAAATTTCTTCAACCCCAATTTGCAAC 180
QY 181 GGAGATGGGAATGTCAATATCATCAATTCAGAGAGAGAGCCGAGAGAGTTCACTACCC 240
DB 181 GGAGATGGGAATGTCAATATCATCAATTCAGAGAGAGAGCCGAGAGAGTTCACTACCC 240
QY 241 ATAAATCCATTTTATAGAGCACTATATTCATGCTTTATATTAATCAACCGAGA 300
DB 241 ATAAATCCATTTTATAGAGCACTATATTCATGCTTTATATTAATCAACCGAGA 300
QY 301 GCAATTTGATCTGAATATCATATTTACTGACACCTTGATATCATTCACAAGAGATACT 360
DB 301 GCAATTTGATCTGAATATCATATTTACTGACACCTTGATATCATTCACAAGAGATACT 360
QY 361 GAAACCAAGACATCTCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 361 GAAACCAAGACATCTCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420

```

```

Db      361  GAAACCAAGACATCTCGAATGAGAGAAAAAATTCTTTCTTACCATTCCTGTCATTACC 420
Qy      421  GAAAGTGAAGAAACCGGATTTAGTTCATGCGAGAGATTAGTGCAGTTTGCCCAACT 480
Db      421  GAAAGTGAAGAAACCGGATTTAGTTCATGCGAGAGATTAGTGCAGTTTGCCCAACT 480
Qy      481  TTATTTATCCCATTTTATCCCAATGTTATCAGTACGAATTAATTTTGGACAACGTT 540
Db      481  TTATTTATCCCATTTTATCCCAATGTTATCAGTACGAATTAATTTTGGACAACGTT 540
Qy      541  GCACAGATTGCAACATTTGTTATGCCCAAAAAAGATAGATCTGGGTTTATGTTGCACT 600
Db      541  GCACAGATTGCAACATTTGTTATGCCCAAAAAAGATAGATCTGGGTTTATGTTGCACT 600
Qy      601  GGAATTTATGCTGATTTGATATGAAAGATTGCGCGAGTTGTTGTAATGACGGTTT 660
Db      601  GGAATTTATGCTGATTTGATATGAAAGATTGCGCGAGTTGTTGTAATGACGGTTT 660
Qy      661  CAGGTTCTAGAAAGTATCTGAGAAAGTTCCCAACAGATTGAAAAAATTTGATGAACT 720
Db      661  CAGGTTCTAGAAAGTATCTGAGAAAGTTCCCAACAGATTGAAAAAATTTGATGAACT 720
Qy      721  AACTGGGAATTCAAACATGAAAGATGTACATTACCAACGSAATCAAGTTATTAATGCGT 780
Db      721  AACTGGGAATTCAAACATGAAAGATGTACATTACCAACGSAATCAAGTTATTAATGCGT 780
Qy      781  GACGTCAAGGGTGGCTCAGAAACACCAAAATGTTGTCAGAGTCTCCAAATGSAAG 840
Db      781  GACGTCAAGGGTGGCTCAGAAACACCAAAATGTTGTCAGAGTCTCCAAATGSAAG 840
Qy      841  GAAAAAGCCAGAAAGAGCTCTGTGTGTATGACACAGCTTAATAGTCCCAATTTACGTT 900
Db      841  GAAAAAGCCAGAAAGAGCTCTGTGTGTATGACACAGCTTAATAGTCCCAATTTACGTT 900
Qy      901  ATGAAAGAAATTAAGGGAATTCGCTGAAAAATACGACTCAGACCCAGAGACTTATATTA 960
Db      901  ATGAAAGAAATTAAGGGAATTCGCTGAAAAATACGACTCAGACCCAGAGACTTATATTA 960
Qy      961  GAGTTAGATCATTCCTGTAGGCTTGAAGTGGATTAAGAACTCAGAAAAAGGTTA 1020
Db      961  GAGTTAGATCATTCCTGTAGGCTTGAAGTGGATTAAGAACTCAGAAAAAGGTTA 1020
Qy      1021  CAAGCATTTAACAACAAAATCAGAGGTTCCAAATGTAACCTGATGTCNAACCCAGTTGTG 1080
Db      1021  CAAGCATTTAACAACAAAATCAGAGGTTCCAAATGTAACCTGATGTCNAACCCAGTTGTG 1080
Qy      1081  GACCGTTGCAAGAGATTCCTGGTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db      1081  GACCGTTGCAAGAGATTCCTGGTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Qy      1141  GATGCAATAGCTGTATTAAGTATGTAATCAAGTGGGAAATTTTAAGCAAGAACTCTT 1200
Db      1141  GATGCAATAGCTGTATTAAGTATGTAATCAAGTGGGAAATTTTAAGCAAGAACTCTT 1200
Qy      1201  GAAAAATCCAGATTTATTTATATATGTTTACTGGGTTTGGAAAGCAACAGAGGT 1260
Db      1201  GAAAAATCCAGATTTATTTATATATGTTTACTGGGTTTGGAAAGCAACAGAGGT 1260
Qy      1261  GTACTTGAAGAAAAACCAAGAGCTATATAGGTTTAA 1299
Db      1261  GTACTTGAAGAAAAACCAAGAGCTATATAGGTTTAA 1299

```

```

RESULT 5
AX087874 547 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 1 from Patent WO0114533.
ACCESSION AX087874
VERSION AX087874.1 GI:13396867
KEYWORDS
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

```

```

REFERENCE
1 Rosamond, J.D. and Schnell, N.F.
Phosphomethyltransferase kinase (pmk) gene (erg8) from candida albicans
Patent: WO 0114533-A 1 01-MAR-2001;
Astrazeneca AB (SE)
FEATURES
source
1..547
Location/Qualifiers
/organism="Candida albicans"
/mol_type="unassigned DNA"
/db_xref="taxon:5476"

```

```

Query Match 32.4%; Score 421; DB 6; Length 547;
Best Local Similarity 99.8%; Pred. No. 5.6e-209;
Matches 471; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      828  CCATGGAAGAAAGAAAGCCAGAAAGAGCTCTGTGTGTATGACACGCTTAATATGTC 887
Db      1  CCATGGAAGAAAGAAAGCCAGAAAGAGCTCTGTGTGTATGACACGCTTAATATGTC 60
Qy      888  CAATTACAGTTTAAAGAGAAATTAAGGAAATGCGTGAATAATGCACTCAGACCCAGA 947
Db      61  CAATTACAGTTTAAAGAGAAATTAAGGAAATGCGTGAATAATGCACTCAGACCCAGA 120
Qy      948  GACTTATATTAAGAGTTAGATCATTCCTGTGAGCTTTGACCTGTTGCGATTAAACAT 1007
Db      121  GACTTATATTAAGAGTTAGATCATTCCTGTGAGCTTTGACCTGTTGCGATTAAACAT 180
Qy      1008  CAGAAAGGGTTTACAGCAATTACACAAAAATCAGAGGTTCCAAATGAACTGATGTCCA 1067
Db      181  CAGAAAGGGTTTACAGCAATTACACAAAAATCAGAGGTTCCAAATGAACTGATGTCCA 240
Qy      1066  AACCCAGTTGTGACCCGTGTCAGAGATTCCTGCTGTGTGTGTGTGTGTGTGTGTGTGT 1127
Db      241  AACCCAGTTGTGACCCGTGTCAGAGATTCCTGCTGTGTGTGTGTGTGTGTGTGTGTGT 300
Qy      1128  TCGTGTGATACAGTACATGCTGTATTAAGTGTGAAATCAAGTGGGAAATTTTAA 1187
Db      301  TCGTGTGATACAGTACATGCTGTATTAAGTGTGAAATCAAGTGGGAAATTTTAA 360
Qy      1188  GCAGAAACCTTTGAAATCCAGATTTATTTATATGTTTACTGCGGTTGATTTGAGAGA 1247
Db      361  GCAGAAACCTTTGAAATCCAGATTTATTTATATGTTTACTGCGGTTGATTTGAGAGA 420
Qy      1248  GCAAACAGAGGTGTACTTGAAGAAACCAAGAAACATATATAGGTTTAA 1299
Db      421  GCAAACAGAGGTGTACTTGAAGAAACCAAGAAACATATATAGGTTTAA 472

```

```

RESULT 6
AX087876/c 577 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 3 from Patent WO0114533.
ACCESSION AX087876
VERSION AX087876.1 GI:13396869
KEYWORDS
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetes; mitosporic Saccharomycetes; Candida.
REFERENCE
1 Rosamond, J.D. and Schnell, N.F.
Phosphomethyltransferase kinase (pmk) gene (erg8) from candida albicans
Patent: WO 0114533-A 3 01-MAR-2001;
Astrazeneca AB (SE)
FEATURES
source
1..577
Location/Qualifiers
/organism="Candida albicans"
/mol_type="unassigned DNA"
/db_xref="taxon:5476"

```

ORIGIN

Query Match 25.7%; Score 334; DB 6; Length 577;
 Best Local Similarity 99.7%; Pred. No. 2.4e-163;
 Matches 384; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 331 GACCCGTGATATCATTCACAGAAGATACGAAACCAAGCATCTTCGAATGAGAAAA 390
 DB 421 GACCCGTGATATCATTCACAGAAGATACGAAACCAAGCATCTTCGAATGAGAAAA 362
 QY 391 ACAATTTCTTTACCAATTCCTGTCGCATTACCGAAGTGGAAAAAGCCGATTAAGTTGTCG 450
 DB 361 ACTTTTCTTACCAATTCCTGTCGCATTACCGAAGTGGAAAAAGCCGATTAAGTTGTCG 302
 QY 451 GCAGGATTAAGTGCAGTTGTCGCACAGATTATATCCCATTTATCCCAATGTTATC 510
 DB 301 GCAGGATTAAGTGCAGTTGTCGCACAGATTATATCCCATTTATCCCAATGTTATC 242
 QY 511 AGTACGAATTAAGATATTTTGGCACAAGTGGACAGATTGACATGTTATGCCCCAAAA 570
 DB 241 AGTACGAATTAAGATATTTTGGCACAAGTGGACAGATTGACATGTTATGCCCCAAAA 182
 QY 571 AGATATGATCTGGGTTTGAATGTTGCACTGCAATTATGCTGATGATATAGAGA 630
 DB 181 AGATATGATCTGGGTTTGAATGTTGCACTGCAATTATGCTGATGATATAGAGA 122
 QY 631 TTTACGCCGCTTGTGATTAATGACGCTTTTCAGTTCTAGAAAGTATCTTGAAGATTC 690
 DB 121 TTTACGCCGCTTGTGATTAATGACGCTTTTCAGTTCTAGAAAGTATCTTGAAGATTC 62
 QY 691 CCCACAGAGTTGAAAAAATTTGATTG 715
 DB 61 CCCACAGAGTTGAAAAAATTTGATTG 37

RESULT 7
 LOCUS AX485723 65 bp DNA linear PAT 16-AUG-2002
 DEFINITION Sequence 3023 from Patent WO02053728.
 ACCESSION AX485723
 VERSION AX485723.1 GI:22319939
 KEYWORDS
 SOURCE Candida albicans
 ORGANISM Candida albicans
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE 1 Roemer, T., Jiang, B., Boone, C., Bussey, H. and Ohlsen, K.L.
 TITLE Gene disruption methodologies for drug target discovery
 JOURNAL Patent: WO 02053728-A 3023 11-JUL-2002;
 Eiltra Pharmaceuticals, Inc. (US)

FEATURES
 source 1..65
 Location/Qualifiers
 1..65
 /organism="Candida albicans"
 /mol_type="unassigned DNA"
 /db_xref="taxon:5476"

ORIGIN

Query Match 5.0%; Score 65; DB 6; Length 65;
 Best Local Similarity 100.0%; Pred. No. 3.6e-22;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTCAAAAGCATTTAGTGAACCTGGAAGACATTTCTGCTGTGATATTTGGTTCTT 60
 DB 65 ATGTCAAAAGCATTTAGTGAACCTGGAAGACATTTCTGCTGTGATATTTGGTTCTT 6
 QY 61 GAGCC 65
 DB 5 GAGCC 1

RESULT 8
 LOCUS AX241110/c 90 bp DNA linear PAT 26-SEP-2001
 DEFINITION Sequence 348 from Patent WO0160975.

ACCESSION AX241110
 VERSION AX241110.1 GI:15797985
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM synthetic construct
 REFERENCE 1 Roemer, T., Jiang, B., Boone, C. and Bussey, H.
 TITLE Gene disruption methodologies for drug target discovery
 JOURNAL Patent: WO 0160975-A 348 23-AUG-2001;
 Eiltra Pharmaceuticals, Inc. (US)

FEATURES
 source 1..90
 Location/Qualifiers
 1..90
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="DNA primer"

ORIGIN

Query Match 5.0%; Score 65; DB 6; Length 90;
 Best Local Similarity 100.0%; Pred. No. 3.6e-22;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTCAAAAGCATTTAGTGAACCTGGAAGACATTTCTGCTGTGATATTTGGTTCTT 60
 DB 65 ATGTCAAAAGCATTTAGTGAACCTGGAAGACATTTCTGCTGTGATATTTGGTTCTT 6
 QY 61 GAGCC 65
 DB 5 GAGCC 1

RESULT 9
 LOCUS AX087877 25 bp DNA linear PAT 17-MAR-2001
 DEFINITION Sequence 4 from Patent WO0114533.
 ACCESSION AX087877
 VERSION AX087877.1 GI:13396870
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM synthetic construct
 REFERENCE 1 Rosamond, J.D. and Schnell, N.F.
 TITLE Phosphomevalonate kinase (pmk) gene (erg8) from candida albicans
 JOURNAL Patent: WO 0114533-A 4 01-MAR-2001;
 Astrazeneca AB (SE)

FEATURES
 source 1..25
 Location/Qualifiers
 1..25
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="Single-stranded oligonucleotide"

ORIGIN

Query Match 1.9%; Score 25; DB 6; Length 25;
 Best Local Similarity 100.0%; Pred. No. 0.37;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 AAGTTATTATCCCATTTATCCCC 501
 DB 25 AAGTTATTATCCCATTTATCCCC 1

RESULT 10
 LOCUS AX087884/c 36 bp DNA linear PAT 17-MAR-2001
 DEFINITION Sequence 11 from Patent WO0114533.
 ACCESSION AX087884
 VERSION AX087884.1 GI:13396876
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM synthetic construct

```

artificial sequences.
REFERENCE
1
AUTHORS
Rosamond J.D. and Schnell N.F.
TITLE
Phosphomevalonate kinase (pmk) gene (erg8) from candida albicans
JOURNAL
Patent: WO 014533-A 11 01-MAR-2001;
Astrazeneca AB (SE)
FEATURES
source
1..36
Location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Single-stranded oligonucleotide"
ORIGIN
Query Match 1.8%; Score 24; DB 6; Length 36;
Best Local Similarity 100.0%; Pred.No.1.2;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db
1276 CCAGAGACTATAGTTTATTA 1299
|||||
36 CCAGAGACTATAGTTTATTA 13
RESULT 11
AE014174.2/c
WPCOMMENT
Sequence split into 4 fragments LOCUS AE014174 Accession AE014174
Fragment Name Begin End
AE014174_0 1 110000
AE014174_1 100001 210000
AE014174_2 200001 310000
AE014174_3 300001 400029
Continuation (3 of 4) of AE014174 from base 200001 (AE014174 Mus musculus plebeald delcti
Query Match 1.8%; Score 23; DB 10; Length 110000;
Best Local Similarity 100.0%; Pred.No.2.6;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db
1211 ATTATTTTATATGTTTACTG 1233
|||||
92633 ATTATTTTATATGTTTACTG 92611
RESULT 12
AC115701
LOCUS
AC115701 Mus musculus chromosome 14 clone RP23-434P22 map 14, WORKING DRAFT
SEQUENCE, 9 ordered pieces.
AC115701
AC115701.4 GI:44886761
HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 201532)
Birren,B., Nuebaum,C. and Lander,B.
Mus musculus chromosome 14, clone RP23-434P22
Unpublished
2 (bases 1 to 201532)
Birren,B., Linton,L., Nuebaum,C., Lander,B., Ali,A., Allen,N.,
Anderson,S., Barina,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouhagalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,I.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Galagan,J., Gardyna,S.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazates,R.,
Lander,B., Lencock,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Menus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,

```

```

Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunhthang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rie,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnpack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Straus,N., Subramanian,A., Talame,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.T., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 201532)
TITLE
JOURNAL
REFERENCE
AUTHORS
Birren,B., Nuebaum,C., Lander,B., Abouelleil,A., Allen,N.,
Anderson,M., Arabchi,H.M., Barina,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Bouhagalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., Deatellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Lander,B., Lencock,J.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunhthang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rie,C., Rogov,P.,
Roman,J., Schauer,S., Schnpack,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-MAR-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2004 this sequence version replaced gi:31455673.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20308
Center clone name: 434_P 22
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 197912 bases at least Q40
Consensus quality: 198811 bases at least Q30
Consensus quality: 199046 bases at least Q20
Insert size: 200000; agarose-fp
Insert size: 199227; sum-of-ctngs
Quality coverage: 10.9 in Q20 bases; agarose-fp
Quality coverage: 10.9 in Q20 bases; sum-of-ctngs
-----
NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 92555: contig of 92555 bp in length
92556 92655: gap of 100 bp
92656 95400: contig of 2745 bp in length
95401 95500: gap of 100 bp

```

```

* 95501 130686: contig of 35186 bp in length
* 130687 130786: gap of 100 bp
* 130787 134173: contig of 3387 bp in length
* 134174 134273: gap of 100 bp
* 134274 166265: contig of 3192 bp in length
* 166266 166365: gap of 100 bp
* 166366 170490: contig of 4125 bp in length
* 170491 170590: gap of 100 bp
* 170591 181579: contig of 10989 bp in length
* 181580 181679: gap of 100 bp
* 181680 183883: contig of 2204 bp in length
* 183884 183983: gap of 100 bp
* 183984 201532: contig of 17549 bp in length.
FEATURES
source
1..201532
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="14"
/map="14"
/clone="RP23-434P22"
/clone_11b="RPC1-23 Female Mouse BAC"
misc_feature
1..92555
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
misc_feature
92556..95400
/note="assembly_fragment"
95501..130686
/note="assembly_fragment"
130787..134173
/note="assembly_fragment"
134274..166265
/note="assembly_fragment"
166366..170490
/note="assembly_fragment"
170591..181579
/note="assembly_fragment"
181680..183883
/note="assembly_fragment"
183984..201532
/note="assembly_fragment"
clone_end:T7
vector_side:right"

```

ORIGIN

Query Match 1.8%; Score 23; DB 2; Length 201532;
 Best Local Similarity 100.0%; Pred. No. 2.5;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1211 ATTATTTTCATATGTTTACTG3 1233
 Db 44001 ATTATTTTCATATGTTTACTG3 44023

```

RESULT 13
AX241232/c
LOCUS
DEFINITION Sequence 470 from Patent WO0160975.
ACCESSION AX241232
VERSION AX241232.1 GI:15798107
KEYWORDS
SOURCE
ORGANISM
synthetic construct
artificial sequences.
REFERENCE
1
AUTHORS Roemer,T., Jiang,B., Boone,C. and Bussey,H.
TITLE Gene disruption methodologies for drug target discovery
JOURNAL Patent: WO 0160975-A 470 23-AUG-2001;
Elitra Pharmaceuticals, Inc. (US)
FEATURES
source
1..22
/organism="synthetic construct"

```

ORIGIN

Query Match 1.7%; Score 22; DB 6; Length 22;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 454 GGATTAGTGTGAGTTGTGCCA 475
 Db 22 GGATTAGTGTGAGTTGTGCCA 1

```

RESULT 14
AX487723/c
LOCUS
DEFINITION Sequence 5023 from Patent WO02053728.
ACCESSION AX487723
VERSION AX487723.1 GI:22321803
KEYWORDS
SOURCE
ORGANISM
Candida albicans
Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
1
AUTHORS Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
TITLE Gene disruption methodologies for drug target discovery
JOURNAL Patent: WO 02053728-A 5023 11-JUL-2002;
Elitra Pharmaceuticals, Inc. (US)
FEATURES
source
1..22
/organism="Candida albicans"
/mol_type="unassigned DNA"
/db_xref="taxon:5476"

```

ORIGIN

Query Match 1.7%; Score 22; DB 6; Length 22;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 454 GGATTAGTGTGAGTTGTGCCA 475
 Db 22 GGATTAGTGTGAGTTGTGCCA 1

```

RESULT 15
AC090350
LOCUS
DEFINITION Homo sapiens chromosome 11 clone RP11-655N17 map 11, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC090350
VERSION AC090350.3 GI:13625484
KEYWORDS
HTG: HTGS PHASE0
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 58879)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP11-655N17
JOURNAL Unpublished
AUTHORS
2 (bases 1 to 58879)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Batra,N., Baetjen,V., Boguslavsky,L., Bouhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choedel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardina,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Lacroque,K., Lamazares,R., Landers,T.,
Lenczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,

```

TITLE
JOURNAL
COMMENT

McPheeters, R., Meldrim, J., Meneus, L., Mhova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Norb, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roettli, M.,
Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P.,
Sounez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Straus, C., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Travers, M., Travis, N., Trigliolo, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J.,
Zemek, L., Zimmer, A. and Zody, W.

Direct Submission

Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 15, 2001 this sequence version replaced gi:13431044.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L12631

Center clone name: 655_N_17

* NOTE: This record contains 70 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 760: contig of 760 bp in length
* 761 860: gap of 100 bp
* 861 1563: contig of 703 bp in length
* 1564 1663: gap of 100 bp
* 1664 2379: contig of 716 bp in length
* 2380 2479: gap of 100 bp
* 2480 3209: contig of 730 bp in length
* 3210 3309: gap of 100 bp
* 3310 4045: contig of 736 bp in length
* 4046 4145: gap of 100 bp
* 4146 4887: contig of 742 bp in length
* 4888 4987: gap of 100 bp
* 4988 5757: contig of 770 bp in length
* 5758 5857: gap of 100 bp
* 5858 6609: contig of 752 bp in length
* 6610 6709: gap of 100 bp
* 6710 7434: contig of 725 bp in length
* 7435 7534: gap of 100 bp
* 7535 8289: contig of 755 bp in length
* 8290 8389: gap of 100 bp
* 8390 9138: contig of 749 bp in length
* 9139 9238: gap of 100 bp
* 9239 9985: contig of 747 bp in length
* 9986 10085: gap of 100 bp
* 10086 10823: contig of 738 bp in length
* 10824 10923: gap of 100 bp
* 10924 11672: contig of 749 bp in length
* 11673 11772: gap of 100 bp
* 11773 12501: contig of 729 bp in length
* 12502 12601: gap of 100 bp
* 12602 13350: contig of 749 bp in length
* 13351 13450: gap of 100 bp
* 13451 14181: contig of 731 bp in length
* 14182 14281: gap of 100 bp
* 14282 15006: contig of 725 bp in length

15007 15106: gap of 100 bp
* 15107 15842: contig of 736 bp in length
* 15843 15942: gap of 100 bp
* 15943 16594: contig of 752 bp in length
* 16595 16794: gap of 100 bp
* 16795 17547: contig of 753 bp in length
* 17548 17647: gap of 100 bp
* 17648 18396: contig of 749 bp in length
* 18397 18496: gap of 100 bp
* 18497 19235: contig of 739 bp in length
* 19236 19335: gap of 100 bp
* 19336 20099: contig of 764 bp in length
* 20100 20199: gap of 100 bp
* 20200 20945: contig of 746 bp in length
* 20946 21045: gap of 100 bp
* 21046 21794: contig of 749 bp in length
* 21795 21894: gap of 100 bp
* 21895 22622: contig of 728 bp in length
* 22623 22722: gap of 100 bp
* 22723 23451: contig of 729 bp in length
* 23452 23551: gap of 100 bp
* 23552 24286: contig of 735 bp in length
* 24287 24386: gap of 100 bp
* 24387 25122: contig of 736 bp in length
* 25123 25222: gap of 100 bp
* 25223 25976: contig of 754 bp in length
* 25977 26076: gap of 100 bp
* 26077 26827: contig of 751 bp in length
* 26828 26927: gap of 100 bp
* 26928 27676: contig of 749 bp in length
* 27677 28522: contig of 746 bp in length
* 28523 28622: gap of 100 bp
* 28623 29373: contig of 751 bp in length
* 29374 29473: gap of 100 bp
* 29474 30268: contig of 795 bp in length
* 30269 31114: contig of 746 bp in length
* 31115 31214: gap of 100 bp
* 31215 31939: contig of 725 bp in length
* 31940 32039: gap of 100 bp
* 32040 32771: contig of 732 bp in length
* 32772 32871: gap of 100 bp
* 32872 33601: contig of 730 bp in length
* 33602 33701: gap of 100 bp
* 33702 34435: contig of 734 bp in length
* 34436 34535: gap of 100 bp
* 34536 35281: contig of 746 bp in length
* 35282 35381: gap of 100 bp
* 35382 36120: contig of 735 bp in length
* 36121 36220: gap of 100 bp
* 36221 36966: contig of 746 bp in length
* 36967 37066: gap of 100 bp
* 37067 37797: contig of 731 bp in length
* 37798 37897: gap of 100 bp
* 37899 38641: contig of 744 bp in length
* 38642 38741: gap of 100 bp
* 38742 39526: contig of 785 bp in length
* 39527 39626: gap of 100 bp
* 39627 40365: contig of 739 bp in length
* 40366 40465: gap of 100 bp
* 40466 41203: contig of 738 bp in length
* 41204 41303: gap of 100 bp
* 41304 42041: contig of 738 bp in length
* 42042 42141: gap of 100 bp
* 42142 42891: contig of 750 bp in length
* 42892 42991: gap of 100 bp
* 42992 43728: contig of 737 bp in length
* 43729 43828: gap of 100 bp
* 43829 44572: contig of 744 bp in length
* 44573 44672: gap of 100 bp
* 44673 45412: contig of 740 bp in length
* 45413 45512: gap of 100 bp

```
*
* 45513 46227: contig of 715 bp in length
* 46228 46327: gap of 100 bp
* 46328 47081: contig of 754 bp in length
* 47082 47181: gap of 100 bp
* 47182 47934: contig of 753 bp in length
* 47935 48034: gap of 100 bp
* 48035 48790: contig of 756 bp in length
* 48791 48890: gap of 100 bp
* 48891 49630: contig of 740 bp in length
* 49631 49730: gap of 100 bp
* 49731 50471: contig of 741 bp in length
* 50472 50572: gap of 100 bp
* 50573 51316: contig of 745 bp in length
* 51317 51416: gap of 100 bp
* 51417 52149: contig of 733 bp in length
* 52150 52249: gap of 100 bp
* 52250 52995: contig of 746 bp in length
* 52996 53095: gap of 100 bp
* 53096 53845: contig of 750 bp in length
* 53846 53945: gap of 100 bp
* 53946 54772: contig of 727 bp in length
* 54773 55521: contig of 749 bp in length
* 55522 55621: gap of 100 bp
* 55622 56364: contig of 743 bp in length
* 56365 56464: gap of 100 bp
* 56465 57190: contig of 726 bp in length
* 57191 57290: gap of 100 bp
* 57291 58033: contig of 743 bp in length

Query Match 1.7%: Score 22; DB 2; length 58879;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1030 ACACAAATCATGAGGTTCCAA 1051
Db 19543 ACACAAATCATGAGGTTCCAA 19564

RESULT 16
AC149142 153598 bp DNA linear HTG 20-MAY-2004
LOCUS Xenopus tropicalis clone CH216-82H3, *** SEQUENCING IN PROGRESS
DEFINITION
AC149142
AC149142
AC149142
AC149142.2 GI:47523990
VERSION HTG; HTGS PHASE1.
KEYWORDS Xenopus tropicalis (Silurana tropicalis)
SOURCE Xenopus tropicalis
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 153598)
DOE Joint Genome Institute.
Unpublished
2 (bases 1 to 153598)
DOE Joint Genome Institute.
Unpublished
3 (bases 1 to 153598)
DOE Joint Genome Institute.
Direct Submission
Submitted (19-MAY-2004) Production Genomics Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
94598-1698, USA
On May 20, 2004 this sequence version replaced gi:47498095.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
```

```
-----
Project Information
Center Project Name: 3599316
Center clone name: CHOR1-216_82H3
-----
Summary Statistics
Consensus quality: 141761 bases at least Q40
Consensus quality: 145817 bases at least Q30
Consensus quality: 148968 bases at least Q20
Estimated insert size: 170000; agarose-ff estimation
Estimated insert size: 150999; sum-of-contigs estimation
Quality coverage: 2.42 in Q20 bases; agarose-ff estimation
Quality coverage: 2.72 in Q20 bases; sum-of-contigs estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2637 2736: contig of 2636 bp in length
2737 2736: gap of unknown length
4668 4768: contig of 1932 bp in length
4769 4768: gap of unknown length
6661 6660: contig of 1892 bp in length
6661 6761: gap of unknown length
8966 9066: contig of 2206 bp in length
9067 9066: gap of unknown length
12223 12223: contig of 3157 bp in length
12224 12323: gap of unknown length
12324 14853: contig of 2530 bp in length
14854 14953: gap of unknown length
14954 17543: contig of 2590 bp in length
17544 17643: gap of unknown length
17644 19766: contig of 2123 bp in length
19767 19866: gap of unknown length
21772 21772: contig of 1906 bp in length
21773 21873: gap of unknown length
21873 25494: contig of 3622 bp in length
25495 25594: gap of unknown length
25595 28992: contig of 3398 bp in length
28993 31780: gap of unknown length
31781 31880: contig of 2688 bp in length
31881 36776: gap of unknown length
36777 36876: gap of 4896 bp in length
36877 39602: contig of 2726 bp in length
39603 39702: gap of unknown length
39703 44268: contig of 4566 bp in length
44269 44368: gap of unknown length
44369 48059: contig of 3651 bp in length
48060 48159: gap of unknown length
53328 53328: contig of 5169 bp in length
53329 53429: gap of unknown length
53429 58675: contig of 5247 bp in length
58676 58775: gap of unknown length
58776 66270: contig of 7495 bp in length
66271 66370: gap of unknown length
66371 73950: contig of 7580 bp in length
73951 74050: gap of unknown length
74051 83527: contig of 9477 bp in length
83528 83627: gap of unknown length
83628 90690: contig of 7063 bp in length
90691 90799: gap of unknown length
90799 99179: contig of 8289 bp in length
99180 109287: gap of unknown length
109288 109387: contig of 10108 bp in length
109388 119271: gap of unknown length
119272 119371: contig of 9884 bp in length
119372 132522: contig of 13151 bp in length
132523 132622: gap of unknown length
```



```

* 132623 153598; contig of 20976 bp in length.
FEATURES
    Source
        Location/Qualifiers
            1..153598
                /organism="Xenopus tropicalis"
                /mol_type="genomic DNA"
                /db_xref="taxon:8364"
                /clone="CH216-82H3"
                /clone_1ib="CHORI-216 Xenopus tropicalis BAC library"
ORIGIN
Query Match 1.7%; Score 22; DB 2; Length 153598;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 569 AAAAGATGATCTGGGTTTGA 590
DB 68052 AAAAGATGATCTGGGTTTGA 68073
RESULT 17
LOCUS AC103846 171133 bp DNA linear PRI 23-JUL-2002
DEFINITION Homo sapiens chromosome 8, clone RP11-645H17, complete sequence.
ACCESSION AC103846
VERSION AC103846.2 GI:21930254
KEYWORDS HIG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 171133)
AUTHORS Birren,B., Nussbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-645H17
REFERENCE 2 (bases 1 to 171133)
AUTHORS Birren,B., Linton,L., Nussbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Baetien,V., Boguslavsky,L., Boukhalter,B.,
Brown,A., Camarata,J., Campolongo,A., Chang,Y., Chazaro,B.,
Chapel,Y., Colangelo,M., Coppolino,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,M., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Maclean,C., Macdonald,P., Major,J., Levine,R., Liu,G.,
McCarthy,M., McKean,P., McKernan,K., McPheters,R., Meldrim,J.,
Menesh,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Notibu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Petersen,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retter,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Strange-Thomson,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Teefaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,X., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembeck,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (32-NOV-2001) Whitehead Institute/MIT Center for Genome
REFERENCE Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS 3 (bases 1 to 171133)
Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Baetien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
Camarata,J., Chang,Y., Chazaro,B., Chapel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Fato,S., Ferreira,P., Fitzerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,M., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Lander,E., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Menesh,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Notibu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,

```

TITLE	JOURNAL	COMMENT
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuppback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Teafaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.		Submitted (23-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On Jul 23, 2002 this sequence version replaced gi:117149731. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center	Center: Whitehead Institute/ MIT Center for Genome Research	
Center code: MIBR	Web site: http://www-seq.wi.mit.edu	
Contact: sequence.submissions@genome.wi.mit.edu		
Project Information	Center project name: L21845	
Center clone name: 645_H_17		
Location/Qualifiers		
1. 171133	/organism="Homo sapiens"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:9606"	
	/chromosome="8"	
	/map="8"	
	/clone="RP11-645H17"	
	/clone_lib="RP11-645H17 Human Male BAC"	
	complement(1008..1061)	
	/rpt_family="MIR3"	
	2938..3004	
	/rpt_family="L2"	
	complement(3035..3319)	
	/rpt_family="Alusg"	
	3325..3353	
	/rpt_family="(T)n"	
	complement(4041..4138)	
	/rpt_family="L1M4"	
	4139..4436	
	/rpt_family="AluJo"	
	complement(4437..4586)	
	/rpt_family="L1M4"	
	complement(4587..4691)	
	/rpt_family="U6"	
	complement(4725..4803)	
	/rpt_family="MIR3"	
	5077..5307	
	/rpt_family="MIR"	
	5510..5804	
	/rpt_family="Alusx"	
	5895..5923	
	/rpt_family="MIR"	
	complement(7310..7475)	
	/rpt_family="Alusg"	
	complement(7476..7768)	
	/rpt_family="Alusg"	
	8818..9095	
	/rpt_family="Aluub"	
	complement(10244..10597)	
	/rpt_family="THE1C"	
	11648..11960	
	/rpt_family="Aluv"	
	12700..12826	
	/rpt_family="(TG)n"	
	13281..13401	
	/rpt_family="MIR"	
	13452..13606	
	/rpt_family="MIR3"	
	14111..14285	
	/rpt_family="MIR"	

```

repeat_region complement(14296..14473)
/rpt_family="AluJo"
repeat_region complement(14474..14780)
/rpt_family="AluSx"
repeat_region complement(14781..14871)
/rpt_family="AluJo"
repeat_region 17534..17721
/rpt_family="MIR3"
repeat_region 18995..19277
/rpt_family="AluSx"
repeat_region 20245..20309
/rpt_family="TTTGA)n"
repeat_region complement(20328..20623)
/rpt_family="AluJo"
repeat_region 23313..23533
/rpt_family="MIR"
repeat_region complement(24402..24488)
/rpt_family="MIR"
repeat_region 24872..24946
/rpt_family="MERSA"
repeat_region 24947..25235
/rpt_family="AluJo"
repeat_region 25236..25346
/rpt_family="MERSA"
repeat_region 26121..26158
/rpt_family="TTTGA)n"
repeat_region 26818..27092
/rpt_family="AluJo"
repeat_region 27112..27353
/rpt_family="AluY"
repeat_region 27505..27525
/rpt_family="AT_rich"
repeat_region 28048..28068
/rpt_family="AT_rich"
repeat_region complement(28075..28322)
/rpt_family="HAL1"
repeat_region complement(28325..28614)
/rpt_family="HAL1"
repeat_region complement(28615..29122)
/rpt_family="MLT2B3"
repeat_region complement(29123..29234)
/rpt_family="HAL1"
repeat_region 29235..29353
/rpt_family="T1MA7"
repeat_region complement(29354..29366)
/rpt_family="HAL1"
repeat_region complement(29367..29648)
/rpt_family="AluSx"
repeat_region complement(29649..30214)
/rpt_family="HAL1"
repeat_region complement(30244..30533)
/rpt_family="AluSx"
repeat_region complement(30606..30761)
/rpt_family="HAL1"
repeat_region complement(30762..30883)
/rpt_family="HAL1"
repeat_region complement(30884..31173)
/rpt_family="AluJo"
repeat_region complement(31174..31202)
/rpt_family="HAL1"
repeat_region complement(31466..31674)
/rpt_family="HAL1"
repeat_region complement(31721..31800)
/rpt_family="MAD1"
repeat_region complement(32006..32161)
/rpt_family="AluSg/x"
repeat_region 34492..34660
/rpt_family="MIR"
repeat_region 34795..34834
/rpt_family="AT_rich"
repeat_region 35099..35270
/rpt_family="MIR"
repeat_region 36132..36429

```

```

repeat_region complement(36679..36947)
/rpt_family="AluJo"
Query Match 1.7%; Score 22; DB 9; Length 171133;
Best Local Similarity 100.0%; Pred.No. 8.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1030 ACACAAATCAGAGCTTCCA 1051
DB 150942 ACACAAATCAGAGCTTCCA 150921

RESULT 18
AC012311
LOCUS
DEFINITION
ACCESSION
AC012311
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC012311 185418 bp DNA linear HTG 04-MAY-2000
Homo sapiens chromosome 16 clone CTC-542P17, WORKING DRAFT
SEQUENCE, 10 unordered pieces.
AC012311
AC012311.4 GI:7690196
HTG: HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 185418)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
2 (bases 1 to 185418)
DOE Joint Genome Institute.
Direct Submission
Submitted (23-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On May 4, 2000 this sequence version replaced gi:7458754.

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 393685
Center clone name: CIT-HSPC_542P17
-----
Summary Statistics
Consensus quality: 171415 bases at least Q40
Consensus quality: 180780 bases at least Q30
Consensus quality: 182015 bases at least Q20
Estimated insert size: 185000; agarose-fp estimation
Estimated insert size: 184518; sum-of-contigs estimation
Quality coverage: 8.44 in Q20 bases; agarose-fp estimation
Quality coverage: 8.44 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2067: contig of 2067 bp in length
2068
2167: gap of unknown length
2168
2168: contig of 2288 bp in length
2169
4455: gap of unknown length
4456
4556: gap of unknown length
4557
7647: contig of 3092 bp in length
7648
7647: gap of unknown length
7648
7747: gap of unknown length
7748
9887: contig of 2150 bp in length
9888
9997: gap of unknown length
9998
14021: contig of 4024 bp in length
14022
14121: gap of unknown length
14122
25375: contig of 11254 bp in length
25376
25475: gap of unknown length
25476
48014: contig of 22539 bp in length
48015
48114: gap of unknown length

```

FEATURES

- * 48115 96080: contig of 47966 bp in length
- * 96081 96180: gap of unknown length
- * 96181 141589: contig of 45409 bp in length
- * 141590 141689: gap of unknown length
- * 141690 185418: contig of 43729 bp in length.

Location/Qualifiers

1.185418

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="16"

/clone="CTC-542P17"

/clone_lib="Caltech human BAC library C"

ORIGIN

Query Match 1.7%; Score 22; DB 2; Length 185418;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

32 CATTCTTCTGCTGATATTT 53
|||||
18093 CATTCTTCTGCTGATATTT 18114

RESULT 19
AC104393/c 188270 bp DNA linear PRI 25-MAY-2002
LOCUS Homo sapiens chromosome 8, clone CTD-3080F16, complete sequence.
DEFINITION AC104393
AC104393.6 GI:21206340
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 188270)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 8, clone CTD-3080F16
Unpublished
2 (bases 1 to 188270)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouhgalter, B., Brown, A., Camarata, J., Chang, J., Campiano, A., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hages, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Labocque, K., Lamazares, R., Landers, T., Lehoczk, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McNeeters, R., Meldrum, J., Menus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Riese, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (08-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 188270)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouhgalter, B., Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hages, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Labocque, K., Lamazares, R., Landers, T., Lehoczk, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McNeeters, R., Meldrum, J., Menus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Riese, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 188270)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouhgalter, B., Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hages, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Labocque, K., Lamazares, R., Landers, T., Lehoczk, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Riese, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (25-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2002 this sequence version replaced gi:20429465.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L22225
Center clone name: 3080_F_16

FEATURES
source location/Qualifiers
1.188270
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="CTD-3080F16"
/clone_lib="CTD2 Human BAC"
repeat_region repeat_family="AT_rich"
277..306
repeat_region repeat_family="AluSq"
653..909
repeat_region repeat_family="AluSq"
1293..1316
repeat_region repeat_family="GC_rich"
4227..4260

```

repeat_region      /rpt_family="(TG)n"
                    4893. .4934
repeat_region      /rpt_family="GA-rich"
                    7660. .7680
repeat_region      /rpt_family="(T)n"
                    complement(9023. .9076)
repeat_region      /rpt_family="MIR3"
                    10952. .11018
repeat_region      /rpt_family="L2"
                    complement(11049. .11194)
repeat_region      /rpt_family="AluSg/x"
                    11204. .11233
repeat_region      /rpt_family="(T)n"
                    complement(11921. .12018)
repeat_region      /rpt_family="L1MA4"
                    12019. .12317
repeat_region      /rpt_family="AluTo"
                    complement(12318. .12467)
repeat_region      /rpt_family="L1MA4"
                    complement(12468. .12572)
repeat_region      /rpt_family="U6"
                    complement(12606. .12684)
repeat_region      /rpt_family="MIR3"
                    12958. .13188
repeat_region      /rpt_family="MIR"
                    13391. .13685
repeat_region      /rpt_family="AluSx"
                    complement(15189. .15353)
repeat_region      /rpt_family="AluSg"
                    complement(15354. .15646)
repeat_region      /rpt_family="AluSg"
                    16696. .16973
repeat_region      /rpt_family="AluDb"
                    complement(18122. .18475)
repeat_region      /rpt_family="THE1C"
                    19526. .19838
repeat_region      /rpt_family="AluY"
                    20578. .20704
repeat_region      /rpt_family="(TG)n"
                    21159. .21279
repeat_region      /rpt_family="MIR"
                    21342. .21484
repeat_region      /rpt_family="MIR3"
                    21989. .22145
repeat_region      /rpt_family="MIR"
                    complement(22173. .22350)
repeat_region      /rpt_family="AluTo"
                    complement(22351. .22657)
repeat_region      /rpt_family="AluSx"
                    complement(22658. .22748)
repeat_region      /rpt_family="AluTo"
                    25411. .25598
repeat_region      /rpt_family="MIR3"
                    26337. .26368
repeat_region      /note="single clone coverage"
                    26337. .26346
repeat_region      /note="<30 qual single clone coverage"
                    26337. .26340
repeat_region      /note="probably TCG"
                    26353. .26359
repeat_region      /note="<30 qual single clone coverage"
                    26379. .26381
repeat_region      /note="probably CCC, possibly CC"
                    26871. .27153
repeat_region      /rpt_family="AluSx"
                    28121. .28185
repeat_region      /rpt_family="(TTTAA)n"
                    complement(28204. .28499)
repeat_region      /rpt_family="AluTo"
                    31189. .31332
repeat_region      /rpt_family="MIR"
                    complement(32278. .32364)
repeat_region      /rpt_family="MIR"

```

```

repeat_region      32748. .32822
repeat_region      /rpt_family="MER5A"
                    32823. .33111
repeat_region      /rpt_family="AluTo"
                    33112. .33222
repeat_region      /rpt_family="MER5A"
                    33997. .34034

Query Match
Best Local Similarity 1.7%; Score 22; DB 9; Length 188270;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1030 ACACAAATCAGAGTTCCAA 1051
Db      158804 ACACAAATCAGAGTTCCAA 158783

RESULT 20
AC007614/c      192347 bp      DNA      linear      PRI 22-MAR-2003
LOCUS      Homo sapiens chromosome 16 clone RP11-491P9, complete sequence.
DEFINITION      AC007614
ACCESSION      AC007614
VERSION      AC007614.7 GI:29029215
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 192347)
AUTHORS      DOE Joint Genome Institute, Stanford Human Genome Center and Los
                    Alamos National Laboratory.
TITLE      Direct Submission
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 192347)
AUTHORS      Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
                    Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
                    Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
                    Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
                    McMurtry,K., Han,C. and Deaven,L.
TITLE      Direct Submission
JOURNAL      Submitted (20-MAR-1999) Center for Human Genome Studies, DOE Joint
                    Genome Institute, Los Alamos National Laboratory, MS M888, Los
                    Alamos, NM 87545, USA
REFERENCE      3 (bases 1 to 192347)
AUTHORS      DOE Joint Genome Institute.
JOURNAL      Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint
                    Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE      4 (bases 1 to 192347)
AUTHORS      DOE Joint Genome Institute, Stanford Human Genome Center and Los
                    Alamos National Laboratory.
TITLE      Direct Submission
JOURNAL      Submitted (18-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
                    Drive, Walnut Creek, CA 94598, USA
REFERENCE      5 (bases 1 to 192347)
AUTHORS      DOE Joint Genome Institute, Stanford Human Genome Center and Los
                    Alamos National Laboratory.
TITLE      Direct Submission
JOURNAL      Submitted (22-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
                    Drive, Walnut Creek, CA 94598, USA
COMMENT      On Mar 18, 2003 this sequence version replaced gi:16596519.
                    Draft Sequence Produced by DOE Joint Genome Institute
                    www.jgi.doe.gov
                    Finishing Completed at Stanford Human Genome Center and Los Alamos
                    National Laboratory
                    www.sbgc.stanford.edu
                    Quality: Phrap Quality >=40 100% of Sequence;
                    Estimated Total Number of Errors is 0.
FEATURES
source
1. 192347
/oranism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```

ORIGIN /chromosome="16"
/clone="RP11-491P9"

Query Match 1.7%; Score 22; DB 9; Length 192347;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CATTTCTGCTGCTGATATT 53
DB 122248 CATTTCTGCTGCTGATATT 122227

RESULT 21
LOCUS AX087875 21 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 2 from Patent WO0114533.
ACCESSION AX087875
VERSION AX087875.1 GI:13396868
KEYWORDS
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Rosamond, J.D. and Schnell, N.F.
TITLE Phosphomevalonate kinase (pmk) gene (erg8) from candida albicans
JOURNAL Patent: WO 0114533-A 2 01-MAR-2001;
AstraZeneca AB (SE)
FEATURES
source Location/Qualifiers
1..21
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Single-stranded oligonucleotide"

ORIGIN

Query Match 1.6%; Score 21; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1129 GCTGTGATACGATGCATA 1149
DB 1 GCTGTGATACGATGCATA 21

RESULT 22
LOCUS CQ661526 313 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 6452 from Patent WO02070737.
ACCESSION CQ661526
VERSION CQ661526.1 GI:42134774
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Liew, C.C., Marshall, W.E. and Zhang, H.
TITLE Compositions and methods relating to osteoarthritis
JOURNAL Patent: WO 02070737-A 6452 12-SEP-2002;
Chondrogene Inc. (CA)
FEATURES
source Location/Qualifiers
1..313
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 1.6%; Score 21; DB 6; Length 313;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGATCCAGTCTGTG 1135

DB 36 GTGTGATCCAGTCTGTG 56

RESULT 23
LOCUS AX396511 479 bp DNA linear PAT 18-MAY-2002
DEFINITION Sequence 726 from Patent WO0212328.
ACCESSION AX396511
VERSION AX396511.1 GI:21067258
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS King, G.E., Meagher, M.J., Xu, J. and Secrist, H.
TITLE Compositions and methods for the therapy and diagnosis of colon cancer
JOURNAL Patent: WO 0212328-A 726 14-FEB-2002;
CORIXA CORPORATION (US)

FEATURES
source Location/Qualifiers
1..479
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 1.6%; Score 21; DB 6; Length 479;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGATCCAGTCTGTG 1135
DB 41 GTGTGATCCAGTCTGTG 61

RESULT 24
LOCUS AX397039 530 bp DNA linear PAT 18-MAY-2002
DEFINITION Sequence 1254 from Patent WO0212328.
ACCESSION AX397039
VERSION AX397039.1 GI:21067786
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS King, G.E., Meagher, M.J., Xu, J. and Secrist, H.
TITLE Compositions and methods for the therapy and diagnosis of colon cancer
JOURNAL Patent: WO 0212328-A 1254 14-FEB-2002;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
1..530
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 1.6%; Score 21; DB 6; Length 530;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGATCCAGTCTGTG 1135
DB 439 GTGTGATCCAGTCTGTG 419

RESULT 25
LOCUS BD020934 771 bp DNA linear PAT 27-AUG-2002

```

DEFINITION Novel gene and novel gene fragment cloned in human neuroblastoma.
ACCESSION BD020934
VERSION BD020934.1 GI:22562110
KEYWORDS JP 2001245671-A/3172.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS 1 (bases 1 to 771)
TITLE Nakagawara,A.
JOURNAL Novel gene and novel gene fragment cloned in human neuroblastoma
PATENT: JP 2001245671-A 3172 11-SEP-2001;
COMMENT CHIBA PREF,HISAMITSU PHARMACEUTICAL CO INC
OS Homo sapiens (human)
PN JP 2001245671-A/3172
PD 11-SEP-2001
PF 07-MAR-2000 JP 2000159195
PI AKIRA NAKAGAWARA
PC C12N15/09,C12Q1/68,G01N33/53,G01N33/566/(C12Q1/68,C12R1.91),
CC C12N15/00
CC Novel gene and novel gene fragment cloned in human CC
neuroblastoma
FH Key
FT source
Location/Qualifiers
1..771
/organism="Homo sapiens (human)"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 1.6%; Score 21; DB 6; Length 771;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1115 GTGTGTTCCAGTGCTGTG 1135
|||
436 GTGTGTTCCAGTGCTGTG 456

RESULT 26
BD100872 771 bp DNA linear PAT 27-AUG-2002
LOCUS Novel genes cloned in humanneuroblastoma and fragments thereof.
DEFINITION BD100872.1 GI:22646446
ACCESSION BD100872
VERSION WO 0166719-A/3172.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 771)
JOURNAL Novel genes cloned in humanneuroblastoma and fragments thereof
PATENT: WO 0166719-A 3172 13-SEP-2001;
COMMENT CHIBA PREF,HISAMITSU PHARMACEUTICAL CO INC,AKIRA NAKAGAWARA
OS Homo sapiens (human)
PN WO 0166719-A/3172
PD 13-SEP-2001
PF 02-MAR-2001 WO 2001JP001629
PR 07-MAR-2000 JP 00P 159195
PI AKIRA NAKAGAWARA
PC C12N15/11,C12Q1/68,G01N33/53,G01N33/566
CC Novel genes cloned in humanneuroblastoma and fragments thereof
FH Key
FT source
Location/Qualifiers
1..771
/organism="Homo sapiens (human)"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source
Location/Qualifiers
1..771
/organism="Homo sapiens (human)"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 1.6%; Score 21; DB 6; Length 771;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1115 GTGTGTTCCAGTGCTGTG 1135
|||
436 GTGTGTTCCAGTGCTGTG 456

RESULT 27
BD0717294 1444 bp DNA linear PAT 03-FEB-2004
LOCUS Sequence 3228 from Patent WO02068579.
DEFINITION BD0717294
ACCESSION BD0717294
VERSION CQ717294.1 GI:42278151
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1
JOURNAL Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
PATENT: WO 02068579-A 3228 06-SEP-2002;
COMMENT Kites, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
PE Corporation (NY) (US)
Location/Qualifiers
1..1444
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 1.6%; Score 21; DB 6; Length 1444;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1115 GTGTGTTCCAGTGCTGTG 1135
|||
1303 GTGTGTTCCAGTGCTGTG 1323

RESULT 28
AF068483 1596 bp mRNA linear MAM 10-JUN-1998
LOCUS Oryctolagus cuniculus chaperonin Cct6 mRNA, complete cds.
DEFINITION AF068483
ACCESSION AF068483
VERSION AF068483.1 GI:3201993
KEYWORDS Oryctolagus cuniculus (rabbit)
SOURCE Oryctolagus cuniculus
ORGANISM Oryctolagus cuniculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
TITLE 1 (bases 1 to 1596)
JOURNAL Schwartz,G.J., Segal,G.B. and Kittelberger,A.M.
Rabbit Chaperonin Cct6 (Tcpx)
Unpublished
REFERENCE 2 (bases 1 to 1596)
AUTHORS Schwartz,G.J., Segal,G.B. and Kittelberger,A.M.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1998) Pediatrics, University of Rochester, 601
Elmwood Ave, Rochester, NY 14642, USA
Location/Qualifiers
1..1596
/organism="Oryctolagus cuniculus"
/mol_type="mRNA"
/strain="New Zealand"
/db_xref="taxon:9986"
/notes="Tcpx"

FEATURES
source
Location/Qualifiers
1..1596
/organism="Oryctolagus cuniculus"
/mol_type="mRNA"
/strain="New Zealand"
/db_xref="taxon:9986"
/notes="Tcpx"

```

/codon_start=1
/product="chaperonin Cct6"
/protein_id="AAC19379.1"
/db_xref="GI:3201994"
/translation="MAAVKTLNPKAEVARAQAALAVNISAAAGLQDVLTNTLGPKGT
KMLVSGAGDILKLTQDGNVILHMOIQHPASLIATVATQDDITGDGTSNVLIIGEL
LKQADLYISEGLHPIITEGFEAKAKALQVEIQVSKREMRRLTIDVARTSLRTKV
HAEADVTEAVVDSILAIKKODEPIDLPEVMEKHEKSETDSTLIRGLVDHGARR
PMKRRVEDAYILTCNVSLSEYKTEVNSGFPKSAEERKVKARERKPIEDVKKIVE
LKKKVCSDSGKGFVIVNKGIDPFSLDALAKGIVALKRARRKMERLTACGVPIN
SLDDINPDLGAGLVETLTSEKFTIEKCNPRSVTLVKGPKHTLTQIKAIR
DGLRAVKNAIDGCVVPGAGAVEVMAEALIKHKSVMKRAQLGVQAFDALIIPKV
LAONGFPIQETLVKIRTEHSESGQLVGVDLNTGSPMVAEVDNVCVKQLHASC
TVIATNIIIVDEIMRAGMSLKG"

ORIGIN

Query Match 1.6%; Score 21; DB 4; Length 1596;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1115 GTGTGTTCCAGTCTGTG 1135
|||||
Db 1217 GTGTGTTCCAGTCTGTG 1237

RESULT 29
HUMHTR3A 1685 bp mRNA linear PRI 31-DEC-1994
LOCUS Human chaperonin-like protein (HTR3) mRNA, complete cds.
DEFINITION M94083
ACCESSION M94083.1 GI:184461
VERSION chaperonin-like protein.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Segel, G.B., Boal, T.R., Cardillo, T.S., Murant, F.G., Lichtman, M.A.
and Sherman, F.
TITLE Isolation of a gene encoding a chaperonin-like protein by
complementation of yeast amino acid transport mutants with human
cDNA
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (13), 6060-6064 (1992)
MEDLINE 92335237
PUBMED 1352881
COMMENT Original source text: Homo sapiens cDNA to mRNA.
FEATURES
source
location/Qualifiers
1..1685
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_type="b-lymphocyte"
1..1685
/gene="HTR3"
162..1331
/gene="HTR3"
/codon_start=1
/product="chaperonin-like protein"
/protein_id="AA58676.1"
/translation="MDRETLIDVARTSLRTKVAELADVTEAVVDSILAIKKODEPI
DLFMIEMKHEKSETDSTLIRGLVDHGARRPMKRRVEDAYILTCNVSLSEYKTEV
NSGFYKSAEERKVKARERKPIEDVKKIVE LKKKVCSDSGKGFVIVNKGIDPFSL
DALAKGIVALKRARRKMERLTACGVPIN SLDDINPDLGAGLVETLTSEKFTIEKCN
PRSVTLVKGPKHTLTQIKAIR DGLRAVKNAIDGCVVPGAGAVEVMAEALIKHKSVMKRA
QLGVQAFDALIIPKV LAONGFPIQETLVKIRTEHSESGQLVGVDLNTGSPMVAEVDNVC
VKQLHASC TVIATNIIIVDEIMRAGMSLKG"

gene
CDS
ORIGIN
Query Match 1.6%; Score 21; DB 9; Length 1685;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1115 GTGTGTTCCAGTCTGTG 1135
|||||
Db 952 GTGTGTTCCAGTCTGTG 972

RESULT 30
HUMTCP20 2010 bp mRNA linear PRI 13-JAN-1995
LOCUS Human chaperonin protein (tcp20) gene complete cds.
DEFINITION L27706
ACCESSION L27706.1 GI:517064
VERSION Chaperonin; TCP20; TRIC; chaperonin-like protein.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Segel, G.B., Boal, T.R., Cardillo, T.S., Murant, F.G., Lichtman, M.A.
and Sherman, F.
TITLE Isolation of a gene encoding a chaperonin-like protein by
complementation of yeast amino acid transport mutants with human
cDNA
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (13), 6060-6064 (1992)
MEDLINE 92335237
PUBMED 1352881
REFERENCE 2 (bases 1 to 2010)
AUTHORS Richard, L.M., Toth, D., Lichtman, M.A., Hatcl, F.U., Sherman, F. and
Segel, G.B.
TCP20, a subunit of the eukaryotic TRIC chaperonin from humans and
yeast
JOURNAL J. Biol. Chem. 269 (28), 18616-18622 (1994)
MEDLINE 94308102
PUBMED 8034610
COMMENT Original
FEATURES
source
location/Qualifiers
1..2010
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..2010
/gene="TCP20"
87..1682
/gene="TCP20"
/codon_start=1
/product="chaperonin-like protein"
/protein_id="AA61061.1"
/translation="MAAVKTLNPKAEVARAQAALAVNISAAAGLQDVLTNTLGPKGT
KMLVSGAGDILKLTQDGNVILHMOIQHPASLIATVATQDDITGDGTSNVLIIGEL
LKQADLYISEGLHPIITEGFEAKAKALQVEIQVSKREMRRLTIDVARTSLRTKV
HAEADVTEAVVDSILAIKKODEPIDLPEVMEKHEKSETDSTLIRGLVDHGARR
PMKRRVEDAYILTCNVSLSEYKTEVNSGFPKSAEERKVKARERKPIEDVKKIVE
LKKKVCSDSGKGFVIVNKGIDPFSLDALAKGIVALKRARRKMERLTACGVPIN
SLDDINPDLGAGLVETLTSEKFTIEKCNPRSVTLVKGPKHTLTQIKAIR
DGLRAVKNAIDGCVVPGAGAVEVMAEALIKHKSVMKRAQLGVQAFDALIIPKV
LAONGFPIQETLVKIRTEHSESGQLVGVDLNTGSPMVAEVDNVCVKQLHASC
TVIATNIIIVDEIMRAGMSLKG"

gene
CDS
ORIGIN
Query Match 1.6%; Score 21; DB 9; Length 2010;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1115 GTGTGTTCCAGTCTGTG 1135
|||||
Db 1303 GTGTGTTCCAGTCTGTG 1323

RESULT 31
AF385084 2562 bp mRNA linear PRI 12-JUN-2001
LOCUS Homo sapiens heat shock protein mRNA, complete cds.
DEFINITION


```

ACCESSION AF385084
VERSION AF385084.1 GI:14348899
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Lee, Y.-K. and Yoo, Y.-D.
JOURNAL 1 (bases 1 to 2562)
TITLE Homo sapiens chaperonin mRNA sequence
REFERENCE 2 (bases 1 to 2562)
AUTHORS Lee, Y.-K. and Yoo, Y.-D.
JOURNAL Direct Submission
TITLE Submitted (21-MAY-2001) Laboratory of Experimental Therapeutics,
Korea Cancer Center Hospital, Nowon-gu, Seoul 139-706, Korea
FEATURES
source
1..2562
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/feature_type="head and neck cancer"
56..1651
/note="chaperonin"
/codon_start=1
/product="heat shock protein"
/protein_id="AAK61354.1"
/db_xref="GI:14348900"
/translation="MAAVKTLNPKAEVARQAALAVNISARGLDVLRITNLPKGTM
KMVSGAGDIKLTQDGNVLLHMOIQHPASLIAYATADDDITGDTTSNTLIIGEL
LKQADYIIEGHPRIITEGFEAKAEKALQLEEVKVSSEMRRETLIDVARTSLRTKY
HAEIADVLTAADVDSIIATIKODEPIDLPMITMEMKHSEDTSLIRGLVLDHGRH
PMKRRVEDAVIILTCNVSILEYKTEVNSGFYKSAERKLVAKERKLEDRVKLTIE
LKRKCGDSDDKGVVYNQKIDIPFSLDALSKGIVALKRKRNMERTLACGVVALN
SFDDLSPLDLAGAVYEYTLGSEKFTLEKCNPRSVTLIKGNPKRTLTQIKAVR
DGRVAVKNAIDDCVPGAGAVVAAEALIKRKSIVKRAQLGVQAFADALILIPKV
LAQNSGFDETLVKIOAESBSGQLVGVDLNTGEPMAAEVGMNDYCVKQLLHSC
TVIATNILLVDEIMRAGMSLKG"

ORIGIN
Query Match 1.6%; Score 21; DB 9; Length 2562;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1115 GTGTGGTTCAGGTGCTGCTG 1135
|||||
Db 1272 GTGTGGTTCAGGTGCTGCTG 1292

RESULT 32
AB063318 2647 bp mRNA linear PRI 09-AUG-2001
LOCUS Homo sapiens MODP-2, MODP-3 mRNA for acute morphine dependence
DEFINITION related protein 2, acute morphine dependence related peptide 3,
complete cds.
ACCESSION AB063318
VERSION AB063318.1 GI:14517631
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Wang, H., Gao, X., Li, L., Wang, B., Huang, Y. and Han, J.
JOURNAL 1
TITLE Homo sapiens chaperonin (MODP) mRNA expressed in SH-SY5Y
neuroblastoma cells
PUBLISHED Only in Database (2001)
2 (bases 1 to 2647)
Wang, H., Gao, X., Li, L., Wang, B., Huang, Y. and Han, J.
JOURNAL Direct Submission
TITLE Submitted (18-JUN-2001) Haoran Wang, Peking University,
Neuroscience Research Institute, 38 Xueyuan Rd., Beijing 100083,
China (E-mail:haorwang@263.net, URL:www.bjmu.edu.cn,

```

```

FEATURES
source
Tel:86-10-62091150, Fax:86-10-62072207)
Location/Qualifiers
1..2647
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="P11.1"
/clone="DD-7-1"
/cell_line="SH-SY5Y"
/cell_type="neuroblastoma"
/country="China:Beijing"
/note="This gene expression was upregulated in the acute
morphine dependence neurons, but not in the chronic
dependence neurons-This gene was primarily cloned by using
DD-PCR"
137..1732
/gene="MODP-2"
137..1732
/gene="MODP-2"
/function="unknown"
/note="similar to Tric chaperonin reported in L27706 and
heat shock protein reported in AF385084"
/evidence=experimental
/codon_start=1
/product="acute morphine
dependence related protein 2"
/protein_id="BAB61032.1"
/db_xref="GI:14517632"
/translation="MAAVKTLNPKAEVARQAALAVNISARGLDVLRITNLPKGTM
KMVSGAGDIKLTQDGNVLLHMOIQHPASLIAYATADDDITGDTTSNTLIIGEL
LKQADYIIEGHPRIITEGFEAKAEKALQLEEVKVSSEMRRETLIDVARTSLRTKY
HAEIADVLTAADVDSIIATIKODEPIDLPMITMEMKHSEDTSLIRGLVLDHGRH
PMKRRVEDAVIILTCNVSILEYKTEVNSGFYKSAERKLVAKERKLEDRVKLTIE
LKRKCGDSDDKGVVYNQKIDIPFSLDALSKGIVALKRKRNMERTLACGVVALN
SFDDLSPLDLAGAVYEYTLGSEKFTLEKCNPRSVTLIKGNPKRTLTQIKAVR
DGRVAVKNAIDDCVPGAGAVVAAEALIKRKSIVKRAQLGVQAFADALILIPKV
LAQNSGFDETLVKIOAESBSGQLVGVDLNTGEPMAAEVGMNDYCVKQLLHSC
TVIATNILLVDEIMRAGMSLKG"
2357..2524
/gene="MODP-3"
2357..2524
/gene="MODP-3"
/note="putative"
/codon_start=1
/evidence=not experimental
/product="acute morphine
dependence related peptide 3"
/protein_id="BAB61033.1"
/db_xref="GI:14517633"
/translation="MMEGVVTKGHGYWKFELVQLYFQASELRONIRKRVSVIFFS
PKLRFEELIYI"
2614..2619
/note="putative"
/evidence=not experimental
2636
/note="putative"
/evidence=not experimental

ORIGIN
Query Match 1.6%; Score 21; DB 9; Length 2647;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1115 GTGTGGTTCAGGTGCTGCTG 1135
|||||
Db 1353 GTGTGGTTCAGGTGCTGCTG 1373

RESULT 33
CO492936 2837 bp DNA linear PAT 30-JAN-2004
LOCUS Sequence 24803 from Patent WO0160860.
ACCESSION CO492936
VERSION CO492936.1 GI:41458555

```

KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS 1
TITLE Schlegel, R., Endege, W.O. and Monahan, J.E.
JOURNAL Genes differentially expressed in human prostate cancer and their use
Patent: WO 0160860-A 24803 23-AUG-2001;
Millennium Predictive Medicine, Inc. (US)
LOCATION/Qualifiers
1. .2837
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 1.6%; Score 21; DB 6; Length 2837;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1115 GTGTGCTTCACGCTGCTGTG 1135
|||||
Db 1307 GTGTGCTTCACGCTGCTGTG 1327

RESULT 34
AC020045/c 23130 bp DNA linear HTG 03-JAN-2000
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
DEFINITION AC020045
AC020045.1 GI:6664852
VERSION HTG; HTGS_PHASE2.
KEYWORDS Drosophila melanogaster (fruit fly)
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 23130)
REFERENCE Adams, M. and Venter, J.C.
AUTHORS Direct Submission
TITLE Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
JOURNAL Rockville, MD, USA
COMMENT This sequence was identified as CDM:10211847 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1. .23130
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN
Query Match 1.6%; Score 21; DB 2; Length 23130;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1259 GTGTACTTGAGAAAACCG 1279
|||||
Db 554 GTGTACTTGAGAAAACCG 534

RESULT 35
AC092101 38875 bp DNA linear PRI 31-JAN-2004
LOCUS Homo sapiens BAC clone RP11-745J15 from 7, complete sequence.
DEFINITION AC092101
AC092101.5 GI:17647102
VERSION HTG.
KEYWORDS

SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS 1
TITLE Hillier, L.W., Fulton, R.S., Fulton, L.A., Graves, T.A., Pepin, K.H.,
Wagner-McPherson, C., Layman, D., Maas, S.P., Jaeger, S., Walker, R.,
Wylie, K., Sekhon, M., Becker, M.C., O'Connell, M.D., Schaller, M.E.,
Fowell, G.A., Delahunty, K.D., Miner, T.L., Nash, W.E., Cordes, M.,
Du, H., Sun, H., Edwards, J., Bradshaw-Cordum, H., All, J., Andrews, S.,
Isak, A., Vanbrunt, A., Nguyen, C., Du, F., Lamer, B., Courtney, L.,
Kalkbush, J., Ozarek, P., Bielicki, L., Scott, K., Holmes, A.,
Harkins, R., Harris, A., Strong, C.M., Hou, S., Tomlinson, C.,
Dauphin-Kohlberg, S., Kozlowski-Reilly, A., Leonard, C.,
Rock, S.M., Tin-Kolam, A.M., Abbott, A., Minx, P., Maupin, R.,
Stromwater, C., Latreille, P., Miller, N., Johnson, D., Murray, J.,
Wessner, J.P., Wendt, M.C., Yang, S.P., Schultze, B.R., Wallis, J.W.,
Spitich, J., Bleier, T.A., Nelson, J.O., Berkowitz, N., Wohldmann, P.E.,
Cook, L.L., Hickenbotham, M.T., Eldred, J., Williams, D., Bedell, J.A.,
Mardis, E.R., Clifton, S.W., Chissole, S.L., Marra, M.A., Raymond, C.,
Haugen, E., Gillett, W., Zhou, Y., James, R., Phelps, K., Iadonoto, S.,
Bubb, K., Simms, E., Levy, R., Clendenning, J., Kaul, R., Kent, W.J.,
Furey, T.S., Baertsch, R.A., Brent, M.R., Keibler, E., Flitck, P.,
Bork, P., Suyama, M., Bailey, J.A., Portnoy, M.E., Torrents, D.,
Chinwalla, A.T., Gish, W.R., Eddy, S.R., McPherson, J.D., Olson, M.V.,
Richler, E.E., Green, E.D., Waterston, R.H. and Wilson, R.K.
The DNA sequence of human chromosome 7
Nature 424 (6945), 157-164 (2003)

TITLE
JOURNAL MEDLINE
PUBMED 12853948
22737999

REFERENCE
AUTHORS 2 (bases 1 to 38875)
TITLE Abbot, A., Nguyen, C. and Spalding, L.
JOURNAL The sequence of Homo sapiens BAC clone RP11-745J15
Unpublished (2001)

REFERENCE
AUTHORS 3 (bases 1 to 38875)
TITLE Waterston, R.H.
JOURNAL Direct Submission
Submitted (20-JUN-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 38875)
Waterston, R.H.
Direct Submission
Submitted (13-DEC-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 38875)
Waterston, R.H.
Direct Submission
Submitted (03-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 38875)
Waterston, R.
Direct Submission
Submitted (31-JAN-2004) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
7 (bases 1 to 38875)
Wilson, R.
Direct Submission
Submitted (31-JAN-2004) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Dec 13, 2001 this sequence version replaced gi:16596665.

COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: saplens@wustl.wustl.edu
----- Summary Statistics

Center project name: H_NH0745J15

gene	gene	complement(24488..26854)	/gene="OSJNBa0091B22.3"	/note="contains Pfam profile: PF00082 subtilase family"
mRNA	mRNA	complement(424488..26854)	/gene="OSJNBa0091B22.3"	
CDS	CDS	complement(24488..26854)	/gene="OSJNBa0091B22.3"	
		/codon_start=1	/product="subtilase family protein"	
		/protein_id="AA081739.1"	/db_xref="GI:50838978"	
		/translation="WSSFTNLMILNLLVLLALSLSPALCYTHAGEAGSVRTENG		
		TPPEPFAARLYRYIVLVDPPEGAATDDGHRMHESFLPGGRMDGADQARIISY		
		EYFEPEFAARLYRYIVLVDPPEGAATDDGHRMHESFLPGGRMDGADQARIISY		
		AGYGGVAVGLDITGVHAHSPFDRGVPPEPAWRSSCAVAARCNMNLGVKSPV		
		DGGGGDDVGGHGTHTAGNTAGNFTVAGSAGGAGCTAGIAGATVAAKYNCSG		
		CDDDVLVLAGPPEAMKDDGVYVSLGKSSSPPEPEDIATAPSAVAKITVCAKN		
		GGPPEPESNDAPWMLITVYVAGSVRSSTRTVLLNGELVQGLAQGNSTSYELH		
		FSERQPKCNELAGIVGDVAGHVLVCSDDPEVDSVSNAMVAGAVLVINTESEGY		
		TTVLVDYEPGVAVQVTVVAGNHIYEVARSSSAGCKPNAITVFDNLTLSVHPATVA		
		SSFSRSPKVAQVYAKPDLVPLAPGLNIIAAMPPIHOGCGGGGLFVYIGSTMAPIA		
		ASGVAAIVKDSRHDPMLPAIKSALITTSDDVAGAGNPILDEHHRATAPFTGAGHP		
		ARAADPGVIVIAVADYVAGIICALLGAGAGITIVRNLSLGGTLDKTKIPEAQANPT		
		ITVPLPRESSSAAPPEPTVNRVTNVTNVPASSTYMKLEIRSLTMKRSPEKLVFSGVG		
		EKKGSVTVSGCGGGGGEVSGSLSVSGKHMSPVAVAPQPIKLS"		
		24761..24800		
repeat_region	repeat_region	/rpe_family="CCG" n	25200..25223	
repeat_region	repeat_region	/rpe_family="CCG" n	26168..26198	
repeat_region	repeat_region	/rpe_family="CCG" n	28808..28839	
gene	gene	/rpe_family="GC-rich"	complement(29554..33847)	
		/gene="OSJNBa0091B22.4"	/note="contains Pfam profile: PF00665 integrase core domain"	
mRNA	mRNA	complement{join(29554..30791,31005..31870,31974..32076,32179..32500,32537..33554,33696..33847)}	/gene="OSJNBa0091B22.4"	
CDS	CDS	complement{join(29554..30791,31005..31870,31974..32076,32179..32500,32537..33554,33696..33847)}		
Query Match	Query Match	1.6%; Score 21; DB 8; Length 115815;		
Best Local Similarity	Best Local Similarity	100.0%; Pred. No. 29;		
Matches	Matches	21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Db	702 GAAAAATTTGATTGAAAGTAA 722			
	80309 GAAAAATTTGATTGAAAGTAA 80329			
RESULT 37	AC012253/c	128524 bp	DNA	linear
LOCUS	AC012253	128524 bp	DNA	linear
DEFINITION	Homo sapiens chromosome 11 clone RP11-19013 map 11, LOW-PASS			
ACCESSION	AC012253			
VERSION	AC012253.3			
KEYWORDS	HTG; HTGS PHASE0.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
TITLE	1 (bases 1 to 128524)			
JOURNAL	Biiren,B., Linton,L., Nusbaum,C. and Lander,E.			
REFERENCE	Homo sapiens chromosome 11, clone RP11-19013			
AUTHORS	Unpublished			
	2 (bases 1 to 128524)			
	Biiren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,			
	Baldwin,J., Batra,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,			
	Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,			
	Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,			

```

TITLE      Direct Submission
JOURNAL    Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT    On Jul 13, 2000 this sequence version number replaced gi:6715953.
           All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence submissions@genome.wi.mit.edu

----- Project Information -----
Center project name: L3734
Center clone name: 19_J_13
-----

* NOTE: This record contains 141 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
*      843      842: contig of 842 bp in length
*      943      942: gap of 100 bp
*      1822     1821: contig of 879 bp in length
*      1922     1921: gap of 100 bp
*      2814     2813: contig of 892 bp in length
*      2914     2913: gap of 100 bp
*      3632     3631: contig of 718 bp in length
*      3732     3731: gap of 100 bp
*      4607     4606: contig of 875 bp in length
*      4707     4706: gap of 100 bp
*      5583     5582: contig of 876 bp in length
*      5683     5682: gap of 100 bp
*      6549     6548: contig of 866 bp in length
*      6649     6648: gap of 100 bp
*      7519     7518: contig of 870 bp in length
*      7619     7618: gap of 100 bp
*      8463     8462: contig of 844 bp in length
*      8563     8562: gap of 100 bp
*      9419     9418: contig of 856 bp in length
*      9518     9518: gap of 100 bp
*      10400    10399: contig of 881 bp in length
*      10500    10499: gap of 100 bp
*      11387    11386: contig of 887 bp in length
*      1187     1186: gap of 100 bp
*      12377    12376: contig of 890 bp in length
*      12477    12476: gap of 100 bp
*      13365    13364: contig of 888 bp in length
*      13665    13664: gap of 100 bp
*      14323    14322: contig of 858 bp in length
*      14423    14422: gap of 100 bp
*      15289    15288: contig of 866 bp in length
*      15389    15388: gap of 100 bp
*      16270    16269: contig of 861 bp in length
*      16370    16369: gap of 100 bp
*      17237    17237: contig of 868 bp in length
*      17337    17337: gap of 100 bp

```

```

Center Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence\_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: J1734
Center clone name: 19_J_13
-----
* NOTE: This record contains 141 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1          842: contig of 842 bp in length
*      843          942: gap of 100 bp
*      943          1821: contig of 875 bp in length
*      1822          1921: gap of 100 bp
*      1922          2813: contig of 892 bp in length
*      2814          2913: gap of 100 bp
*      2914          3631: contig of 718 bp in length
*      3632          3731: gap of 100 bp
*      3732          4606: contig of 875 bp in length
*      4607          4706: gap of 100 bp
*      4707          5582: contig of 876 bp in length
*      5583          5682: gap of 100 bp
*      5683          6548: contig of 866 bp in length
*      6549          6648: gap of 100 bp
*      6649          7518: contig of 870 bp in length
*      7519          7618: gap of 100 bp
*      7619          8462: contig of 844 bp in length
*      8463          8562: gap of 100 bp
*      8563          9418: contig of 856 bp in length
*      9419          9518: gap of 100 bp
*      9519          10399: contig of 881 bp in length
*      10400          10499: gap of 100 bp
*      10500          11386: contig of 887 bp in length
*      11387          11486: gap of 100 bp
*      11487          12376: contig of 890 bp in length
*      12377          12476: gap of 100 bp
*      12477          13364: contig of 888 bp in length
*      13365          13464: gap of 100 bp
*      13465          14322: contig of 858 bp in length
*      14323          14422: gap of 100 bp
*      14423          15288: contig of 866 bp in length
*      15289          15388: gap of 100 bp
*      15389          16269: contig of 881 bp in length
*      16270          16369: gap of 100 bp
*      16370          17237: contig of 868 bp in length
*      17238          17337: gap of 100 bp

```

```

17338 18216: contig of 879 bp in length
18317 18316: gap of 100 bp
18317 19199: contig of 883 bp in length
19200 19299: gap of 100 bp
19300 20116: contig of 817 bp in length
20117 20216: gap of 100 bp
20217 21072: contig of 856 bp in length
21073 21172: gap of 100 bp
21173 22031: contig of 859 bp in length
22032 22131: gap of 100 bp
22132 23035: contig of 904 bp in length
23036 23135: gap of 100 bp
23136 23995: contig of 860 bp in length
23996 24095: gap of 100 bp
24096 24977: contig of 882 bp in length
24978 25077: gap of 100 bp
25078 25959: contig of 882 bp in length
25960 26059: gap of 100 bp
26060 26928: contig of 869 bp in length
26929 27028: gap of 100 bp
27029 27694: contig of 866 bp in length
27695 27994: gap of 100 bp
27995 28882: contig of 888 bp in length
28883 29828: gap of 100 bp
29829 29928: contig of 846 bp in length
29929 30811: gap of 100 bp
30812 30911: gap of 100 bp
30912 31795: contig of 884 bp in length
31796 31895: gap of 100 bp
31896 32793: contig of 898 bp in length
32794 32893: gap of 100 bp
32894 33795: contig of 902 bp in length
33796 33895: gap of 100 bp
33896 34788: contig of 893 bp in length
34789 34888: gap of 100 bp
34889 35756: contig of 868 bp in length
35757 35856: gap of 100 bp
35857 36786: contig of 930 bp in length
36787 36886: gap of 100 bp
36887 37768: contig of 882 bp in length
37769 37868: gap of 100 bp
37869 38756: contig of 888 bp in length
38757 38856: gap of 100 bp
38857 39744: contig of 888 bp in length
39745 39844: gap of 100 bp
39845 40709: contig of 865 bp in length
40710 40809: gap of 100 bp
40810 41659: contig of 860 bp in length
41670 41769: gap of 100 bp
41770 42656: contig of 887 bp in length
42657 42756: gap of 100 bp
42757 43647: contig of 891 bp in length
43648 43747: gap of 100 bp
43748 44577: contig of 830 bp in length
44578 44677: gap of 100 bp
44679 45544: contig of 867 bp in length
45545 45644: gap of 100 bp
45645 46512: contig of 868 bp in length
46513 46612: gap of 100 bp
46613 47481: contig of 869 bp in length
47482 47581: gap of 100 bp
47582 48437: contig of 856 bp in length
48438 48537: gap of 100 bp
48539 49397: contig of 860 bp in length
49398 49497: gap of 100 bp
49498 50358: contig of 861 bp in length
50359 50458: gap of 100 bp
50459 51336: contig of 878 bp in length
51337 51436: gap of 100 bp
51437 52320: contig of 884 bp in length
52321 52420: gap of 100 bp
52421 53323: contig of 903 bp in length

```

```

53424 53423: gap of 100 bp
53424 54314: contig of 891 bp in length
54315 54414: gap of 100 bp
54415 55295: contig of 881 bp in length
55296 55395: gap of 100 bp
55396 56269: contig of 874 bp in length
56269 56369: gap of 100 bp
56370 57243: contig of 874 bp in length
57244 57343: gap of 100 bp
57344 58214: contig of 871 bp in length
58215 58314: gap of 100 bp
58315 59198: contig of 884 bp in length
59199 59298: gap of 100 bp
59299 60131: contig of 833 bp in length
60132 60231: gap of 100 bp
60232 61118: contig of 887 bp in length
61119 61218: gap of 100 bp
61219 62080: contig of 862 bp in length
62081 62180: gap of 100 bp
62181 63049: contig of 869 bp in length
63050 63149: gap of 100 bp
63150 64055: contig of 908 bp in length
64056 64155: gap of 100 bp
64156 65060: contig of 905 bp in length
65061 65160: gap of 100 bp
65161 66010: contig of 850 bp in length
66011 66110: gap of 100 bp
66111 66984: contig of 874 bp in length
66985 67084: gap of 100 bp
67085 67967: contig of 883 bp in length
67968 68067: gap of 100 bp
68068 68958: contig of 891 bp in length
69058 69058: gap of 100 bp
69059 69923: contig of 865 bp in length
69924 70023: gap of 100 bp

```

Query Match 1.6%; Score 21; DB 2; Length 128524;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1260 TGTACTTGAGAGAAAACCGA 1280

DB 122485 TGTACTTGAGAGAAAACCGA 122465

RESULT 38

AL772361

LOCUS Mouse DNA sequence from clone RP23-136C22 on chromosome 2, complete
 DEFINITION sequence.

ACCESSION AL772361 GI:23380941

VERSION AL772361.5

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE

AUTHORS Tracey, A.

TITLE Direct Submission

JOURNAL Submitted (27-SEP-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgehire, CB10 1SA, UK. E-mail enquiries:

humquerry@sanger.ac.uk

On Sep 30, 2002 this sequence version replaced gi:22798220.

Center: Wellcome Trust Sanger Institute

Genome Center

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.

where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WormPep; Information on the WormPep database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-136C22 is from the RPCI-23 Mouse PAC library constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6.

FEATURES
SOURCE
Location/Qualifiers

1..128685
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-136C22"
/clone_lib="RPCI-23"

ORIGIN

Query Match 1.6%; Score 21; DB 10; Length 128685;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 768 GTTATTAAATGGGACGCTCA 788
Db 34054 GTTATTAAATGGGACGCTCA 34074

RESULT 39
AC096845/c 134105 bp DNA linear VRT 01-OCT-2002
LOCUS Takifugu rubripes clone 214014, complete sequence.
AC096845
VERSION AC096845.2 GI:17386262
KEYWORDS HTG.
SOURCE Takifugu rubripes (Fugu rubripes)
ORGANISM Takifugu rubripes

REFERENCE
AUTHORS Akhter, N., Ayele, K., Beckstrom-Sternberg, S. M., Benjamin, B., Blakesley, R. W., Bouffard, G. G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N. L., Granite, S., Guan, X., Gupta, D., Haghighi, P., Ho, S.-L., Idol, J. R., Karlins, E., Latic, P., Lee-Lin, S.-O., Legaspi, R., Maduro, Q. L., Maduro, V. B., Masello, C., Mastrian, S. D., McCluskey, J. C., McDowell, J., Pearson, P. J., Prasad, A., Sherchenko, Y., Statulic, S., Thomas, J. W., Thomas, P. J., Touchman, J. W., Tsaurgeon, C., Voigt, J. L., Walker, M. A., Wetherby, K. D., Wiggins, L., Zhang, L.-H. and Green, E. D.

TITLE NISC Comparative Sequencing Initiative
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 134105)

AUTHORS Green, E. D.

TITLE Direct Submission

JOURNAL Submitted (28-SEP-2001) NIH Intramural Sequencing Center, 8717
Groveom Circle, Gaithersburg, MD 20877, USA

REFERENCE 3 (bases 1 to 134105)

AUTHORS Green, E. D.

TITLE Direct Submission
JOURNAL Submitted (06-DEC-2001) NIH Intramural Sequencing Center, 8717
Groveom Circle, Gaithersburg, MD 20877, USA

REFERENCE 4 (bases 1 to 134105)
AUTHORS Green, E. D.
TITLE Direct Submission
JOURNAL Submitted (01-OCT-2002) NIH Intramural Sequencing Center, 8717
Groveom Circle, Gaithersburg, MD 20877, USA
On Dec 6, 2001 this sequence version replaced gi:15799588.

----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc.zoo@hgrl.nih.gov
----- Project Information
Center project name: ari
Center clone name: 214014

This sequence was finished as follows unless otherwise noted:
all regions were double-stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of this clone unless otherwise noted. If there are overlapping clones, the overlaps are noted in the beginning and end of the Features section.

FEATURES
SOURCE
Location/Qualifiers

1..134105
/organism="Takifugu rubripes"
/mol_type="genomic DNA"
/db_xref="taxon:31033"
/clone="214014"
/clone_lib="Incyte Genomics"
115823..115825
/note="low quality single stranded/single chemistry region"

ORIGIN

Query Match 1.6%; Score 21; DB 5; Length 134105;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1120 GTTCCAGTCTGCTGATAC 1140
Db 100041 GTTCCAGTCTGCTGATAC 100021

RESULT 40
AL732425/c 145042 bp DNA linear ROD 19-SEP-2002
LOCUS Mouse DNA sequence from clone RP23-36411 on chromosome X, complete sequence.
AL732425
VERSION AL732425.7 GI:23304648
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Akhter, N., Ayele, K., Beckstrom-Sternberg, S. M., Benjamin, B., Blakesley, R. W., Bouffard, G. G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N. L., Granite, S., Guan, X., Gupta, D., Haghighi, P., Ho, S.-L., Idol, J. R., Karlins, E., Latic, P., Lee-Lin, S.-O., Legaspi, R., Maduro, Q. L., Maduro, V. B., Masello, C., Mastrian, S. D., McCluskey, J. C., McDowell, J., Pearson, P. J., Prasad, A., Sherchenko, Y., Statulic, S., Thomas, J. W., Thomas, P. J., Touchman, J. W., Tsaurgeon, C., Voigt, J. L., Walker, M. A., Wetherby, K. D., Wiggins, L., Zhang, L.-H. and Green, E. D.

TITLE NISC Comparative Sequencing Initiative

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 145042)
AUTHORS Philimore, B.
TITLE Direct Submission
JOURNAL Submitted (18-SEP-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk clone requests: clonerequest@sanger.ac.uk

On Sep 23, 2002 this sequence version replaced gi:21711885.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-364L1 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.choil.org/bacpac/home.htm> VECTOR: pBACe3.6.

FEATURES

source

Location/Qualifiers
 1..145042
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP23-364L1"
 /clone_lib="RPCI-23"

ORIGIN

Query Match 1.6%; Score 21; DB 10; Length 145042;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1224 TGTCTACTGGGTGATTGGA 1244
 ||||||||||||||||||
 DB 25558 TGTCTACTGGGTGATTGGA 25538

Search completed: January 26, 2005, 09:09:50
 Job time : 3938 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 26, 2005, 06:38:50 : Search time 484 Seconds
(without alignments)
14088.839 Million cell updates/sec

Title: US-10-069-062-6
Perfect score: 1299
Sequence: 1 atgcctaaagcatcttagtcgc.....aagactatcataggtttataa 1299

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4134886 seqs, 2624710521 residues

Word size : 0
Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : N_Geneseq_23Sep04.*

1: geneeqn19808.*
2: geneeqn19908.*
3: geneeqn20008.*
4: geneeqn20018.*
5: geneeqn20018.*
6: geneeqn20028.*
7: geneeqn20028.*
8: geneeqn20038.*
9: geneeqn20038.*
10: geneeqn20038.*
11: geneeqn20038.*
12: geneeqn20048.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1299	100.0	1299	4 AAD02792	Aad02792 Candida a
2	1299	100.0	1763	4 AAD02791	Aad02791 Candida a
3	891	68.6	1299	4 AAS23422	Aas23422 Candida a
4	891	68.6	1299	6 AB231736	Ab231736 Candida a
5	421	32.4	547	4 AAD02787	Aad02787 S. cerevi
6	324	25.7	577	4 AAD02788	Aad02788 S. cerevi
7	65	5.0	65	6 AB228940	Ab228940 Candida g
8	65	5.0	90	4 AAS23667	Aas23667 Tetraacycl
9	25	1.9	25	4 AAD02790	Aad02790 Candida a
10	24	1.8	36	4 AAD02796	Aad02796 Candida a
11	22	1.7	22	4 AAS23789	Aas23789 Primer B
12	22	1.7	22	6 AB230804	Ab230804 Candida a
13	21	1.6	21	4 AAD02789	Aad02789 Candida a
14	21	1.6	476	9 ACH30537	Ach30537 Human tes
15	21	1.6	479	6 ABK45175	Abk45175 CDNA enco
16	21	1.6	518	6 ABV86952	Abv86952 Human col
17	21	1.6	530	6 ABK45703	Abk45703 CDNA enco
18	21	1.6	771	4 AAI97097	Aai97097 Human neu
19	21	1.6	1574	4 ABQ54714	Abq54714 Human ova
20	21	1.6	1825	12 AAD22869	Aad22869 Human sof
21	21	1.6	2010	6 ABV78079	Abv78079 Hypoxia-r

22	21	1.6	2461	10 ADC29939	Adc29939 Human nov
23	21	1.6	2562	12 ADL82850	Adl82850 Human PRO
24	21	1.6	2562	12 AD019167	Ad019167 Human PRO
25	21	1.6	2562	12 AD060113	Ad060113 Human CCT
26	21	1.6	2562	12 AD060114	Ad060114 Human CCT
27	21	1.6	2594	3 AAI5841	Aai5841 Human pro
28	21	1.6	2647	12 AD060115	Ad060115 Human CCT
29	21	1.6	2637	5 ABV24814	Abv24814 Human pro
30	21	1.6	16682	4 AK79798	Ak79798 Human imm
31	21	1.6	16682	4 AK70010	Ak70010 Human imm
32	21	1.6	21436	4 AK79799	Ak79799 Human imm
33	21	1.6	21436	4 AK70011	Ak70011 Human imm
34	20	1.5	33	4 AAD02795	Aad02795 Candida a
35	20	1.5	502	12 ACH70211	Ach70211 Human gen
36	20	1.5	541	4 AAI81760	Aai81760 Human pol
37	20	1.5	630	6 AB212786	Ab212786 Arabidops
38	20	1.5	46852	8 ABQ76676	Abq76676 Androgen
39	20	1.5	218336	8 ABQ76678	Abq76678 Androgen
40	19	1.5	355	6 AB188534	Ab188534 C maxmore
41	19	1.5	593	8 ACA04641	ACA04641 CDNA enco
42	19	1.5	995	6 ABQ15523	Abq15523 Oligonuc
43	19	1.5	995	6 ABQ15522	Abq15522 Oligonuc
44	19	1.5	1282	5 ABV23937	Abv23937 Human pro
45	19	1.5	1282	5 ABV29820	Abv29820 Human pro
46	19	1.5	4699	10 ADE54957	Ad54957 Human gen
47	19	1.5	4699	10 ADE54973	Ad54973 Human gen
48	19	1.5	4699	10 ADE54965	Ad54965 Human gen
49	19	1.5	4699	10 ADE59402	Ad59402 Human gen
50	19	1.5	4699	10 ADE54953	Ad54953 Human gen
51	19	1.5	4699	10 ADE61072	Ad61072 Human gen
52	19	1.5	4699	10 ADE54961	Ad54961 Human gen
53	19	1.5	4699	10 ADE54969	Ad54969 Human gen
54	19	1.5	5760	4 AAS32472	Aas32472 Human gen
55	19	1.5	5763	4 AAS32471	Aas32471 Human gen
56	19	1.5	6733	4 AB128838	Ab128838 Drosophi
57	19	1.5	10769	12 ADP79308	Adp79308 Human muc
58	19	1.5	12181	4 AB128830	Ab128830 Drosophi
59	19	1.5	96587	9 ADA02501	Ada02501 Mouse Bac
60	19	1.5	96587	10 ADB72239	Ad72239 Mouse Bac
61	19	1.5	96587	10 ADB95749	Ad95749 Mouse Bac
62	19	1.5	110000	6 ABA90521_01	Ab90521_01
63	18	1.4	107	8 AB279966	Ab279966 Potexvitu
64	18	1.4	189	2 AAV75985	Aav75985 Scaphyloc
65	18	1.4	189	8 AB279967	Ab279967 Potexvitu
66	18	1.4	266	4 AAK35148	Aak35148 Human col
67	18	1.4	315	4 AAK55949	Aak55949 Human imm
68	18	1.4	335	5 AAI99252	Aai99252 Human exc
69	18	1.4	335	5 ABA16938	Abal6938 Human ner
70	18	1.4	335	5 ABA16937	Abal6937 Human ner
71	18	1.4	335	5 AAI63602	Aai63602 Human kid
72	18	1.4	336	8 AB279968	Ab279968 Potexvitu
73	18	1.4	345	5 ABA12118	Abal2118 Human ner
74	18	1.4	351	2 AAV78389	Aav78389 Scaphyloc
75	18	1.4	400	2 AAV78244	Aav78244 Scaphyloc
76	18	1.4	400	2 AAV78244	Aav78244 Scaphyloc
77	18	1.4	439	10 AAD32786	Ad32786 Human mit
78	18	1.4	455	6 ABL79441	Ab179441 Human ova
79	18	1.4	472	6 ABL63885	Ab163885 Breast ca
80	18	1.4	494	4 AAI12422	Aai12422 Probe #23
81	18	1.4	494	4 ABA54130	Abas4130 Human foe
82	18	1.4	494	4 AAI33779	Aai33779 Probe #24
83	18	1.4	494	4 ABA43674	Abas3674 Human bre
84	18	1.4	494	4 ABA23877	Abas23877 Probe #23
85	18	1.4	494	4 AAK27844	Aak27844 Human bon
86	18	1.4	494	4 AAK02400	Aak02400 Human bra
87	18	1.4	494	4 ABA527425	Abas27425 Human liv
88	18	1.4	494	5 AAI02337	Aai02337 Probe #23
89	18	1.4	494	6 ABA502298	Abas02298 Human gen
90	18	1.4	508	8 ABV57150	Abv57150 Human pro
91	18	1.4	603	5 AAV714693	Aav714693 Aspergill
92	18	1.4	619	5 AAS72296	Aas72296 DNA encod
93	18	1.4	624	6 ABA69044	Abn69044 Streptoco
94	18	1.4	624	6 ABA69044	Abn69044 Streptoco


```

C 971 16 1.2 645 4 ABK42395 Abk42395 Genomic s
C 972 16 1.2 645 4 ABK42396 Abk42396 Genomic s
C 973 16 1.2 645 9 ADB60552 Adb60552 Connectiv
C 974 16 1.2 645 9 ADB60551 Adb60551 Connectiv
C 975 16 1.2 646 12 ADN13004 Adn13004 Human pro
C 976 16 1.2 655 6 AAS94961 Aas94961 Human DNA
C 977 16 1.2 658 8 ACA22643 Aca22643 Prokaryot
C 978 16 1.2 663 8 ACA29262 Aca29262 Prokaryot
C 979 16 1.2 665 3 AAF14172 Aaf14172 Aspergilli
C 980 16 1.2 673 5 AAd64776 Aad64776 Human sec
C 981 16 1.2 676 5 AAd63680 Aad63680 Human ova
C 982 16 1.2 677 4 AAd10785 Aad10785 Drosophili
C 983 16 1.2 677 5 AAd172909 Aad172909 Human ova
C 984 16 1.2 677 5 AAd138047 Aad138047 Human ova
C 985 16 1.2 677 5 AAd138047 Aad138047 Human ova
C 986 16 1.2 677 5 AAd138047 Aad138047 Human ova
C 987 16 1.2 681 10 AAdC77295 Adc77295 DNA homol
C 988 16 1.2 681 10 AAdC76883 Adc76883 DNA homol
C 989 16 1.2 681 10 AAdK57182 Adk57182 Plant DNA
C 990 16 1.2 681 10 AAdK59429 Adk59429 Plant DNA
C 991 16 1.2 685 8 AAdA26438 Aad26438 Human sec
C 992 16 1.2 685 8 AAdA39792 Aad39792 Human sec
C 993 16 1.2 685 8 AAdC50431 Acc50431 Human sec
C 994 16 1.2 685 10 AAdC73490 Adc73490 Human sec
C 995 16 1.2 685 10 AAdA5981 Aad5981 Gene enco
C 996 16 1.2 685 12 AAd171499 Adl171499 Novel hum
C 997 16 1.2 711 2 AAd215346 Aad215346 Human gen
C 998 16 1.2 711 2 AAdF13451 Aaf13451 Aspergilli
C 999 16 1.2 712 12 AAdJ3482 Adj3482 plant cDN
1000 16 714 5 AAs66691 Aas66691 DNA encod

```

ALIGNMENTS

```

RESULT 1
AAd02792
ID AAd02792 standard; DNA; 1299 BP.
XX
AC AAd02792;
XX
DT 31-MAY-2001 (first entry)
XX
DE Candida albicans phosphomevalonate kinase (PMK) coding sequence.
XX
KM Phosphomevalonate kinase; PMK; ERG8; anti-fungal agent; diagnosis;
XX
OS infection; ds.
XX
XX Candida albicans.
XX
OS
XX
FH Key
FT 1.1299 Location/Qualifiers
FT CDS /tag= a
FT /product= "C. albicans phosphomevalonate kinase (PMK)"
FT /transl_except= (pos:730..732, aa:Glu)
XX
XX WO200114533-A2.
XX
XX 01-MAR-2001.
XX
XX 15-AUG-2000; 2000MO-GB003100.
XX
XX 21-AUG-1999; 99GB-00019766.
XX
XX (ASTR ) ASTRAZENECA AB.
XX (ASTR ) ASTRAZENECA UK LTD.
XX
XX Rosamond JDC, Schnell NF;
XX
XX WPI; 2001-218441/22.
XX P-PSDB; AAY72679.
XX
XX New polypeptides and polynucleotides (ERG8) from Candida albicans, useful
PT

```

```

PT in assays for identifying inhibitors of phosphomevalonate kinase activity
PT and as reagents for diagnosing C. albicans infection.
XX
XX Claim 6; Page 26; 29pp; English.
XX
XX The present sequence is phosphomevalonate kinase (PMK; ERG8) coding
XX sequence from Candida albicans. The ERG8 protein is useful in an assay
XX for identifying compounds that inhibit phosphomevalonate kinase (PMK)
XX activity. These inhibitors are useful as anti-fungal agents. The ERG8 DNA
XX and protein are also useful as reagents for diagnosing C. albicans
XX infection
XX
XX Sequence 1299 BP; 439 A; 220 C; 273 G; 367 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 1299; DB 4; Length 1299;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 ATGTCAAAAGCATTTAGTGCACCTGGAAAAGCATTTCTGCTGTGATATTGGTCTT 60
XX 1 ATGTCAAAAGCATTTAGTGCACCTGGAAAAGCATTTCTGCTGTGATATTGGTCTT 60
XX
XX 61 GAGCCAAATTTATGATGCTTATGTGACAGCATTTGTCATCAGAAATGACGATTATA 120
XX 61 GAGCCAAATTTATGATGCTTATGTGACAGCATTTGTCATCAGAAATGACGATTATA 120
XX
XX 61 GAGCCAAATTTATGATGCTTATGTGACAGCATTTGTCATCAGAAATGACGATTATA 120
XX 61 GAGCCAAATTTATGATGCTTATGTGACAGCATTTGTCATCAGAAATGACGATTATA 120
XX
XX 121 CCAAAAGGACCGATTTTAAAGAAATCTGAATCAAAATTTCTGACCCCAATTTGCAAC 180
XX 121 CCAAAAGGACCGATTTTAAAGAAATCTGAATCAAAATTTCTGACCCCAATTTGCAAC 180
XX
XX 121 CCAAAAGGACCGATTTTAAAGAAATCTGAATCAAAATTTCTGACCCCAATTTGCAAC 180
XX 121 CCAAAAGGACCGATTTTAAAGAAATCTGAATCAAAATTTCTGACCCCAATTTGCAAC 180
XX
XX 181 GGAGATGGAATATATCATATCATCAATCAATCAATCAATCAATCAATCAATCAATCA 240
XX 181 GGAGATGGAATATATCATATCATCAATCAATCAATCAATCAATCAATCAATCAATCA 240
XX
XX 181 GGAGATGGAATATATCATATCATCAATCAATCAATCAATCAATCAATCAATCAATCA 240
XX 181 GGAGATGGAATATATCATATCATCAATCAATCAATCAATCAATCAATCAATCAATCA 240
XX
XX 241 ATAAATCCATTTTATGAGGCACTATATATCATGCTTTTATGCTTATATTCACCGAC 300
XX 241 ATAAATCCATTTTATGAGGCACTATATATCATGCTTTTATGCTTATATTCACCGAC 300
XX
XX 301 GCATTTGATCTTGAATATCATATTTACTGAGACCTCGGATATATTCACCAAGAAAT 360
XX 301 GCATTTGATCTTGAATATCATATTTACTGAGACCTCGGATATATTCACCAAGAAAT 360
XX
XX 301 GCATTTGATCTTGAATATCATATTTACTGAGACCTCGGATATATTCACCAAGAAAT 360
XX 301 GCATTTGATCTTGAATATCATATTTACTGAGACCTCGGATATATTCACCAAGAAAT 360
XX
XX 361 GAAACCAAGACATCTCGAATGAGAAAACATTTCTTACCATCTGCTGATTTAC 420
XX 361 GAAACCAAGACATCTCGAATGAGAAAACATTTCTTACCATCTGCTGATTTAC 420
XX
XX 361 GAAACCAAGACATCTCGAATGAGAAAACATTTCTTACCATCTGCTGATTTAC 420
XX 361 GAAACCAAGACATCTCGAATGAGAAAACATTTCTTACCATCTGCTGATTTAC 420
XX
XX 421 GAACTGGAAGAAAGACCGGATAGTTGTCATGCGGAGAAATGATGTCAGTTGG 480
XX 421 GAACTGGAAGAAAGACCGGATAGTTGTCATGCGGAGAAATGATGTCAGTTGG 480
XX
XX 421 GAACTGGAAGAAAGACCGGATAGTTGTCATGCGGAGAAATGATGTCAGTTGG 480
XX 421 GAACTGGAAGAAAGACCGGATAGTTGTCATGCGGAGAAATGATGTCAGTTGG 480
XX
XX 481 TTATTTATCCATTTTATCCCAATGTTATCAGTAGAATTAAGATTTTGGCAACG 540
XX 481 TTATTTATCCATTTTATCCCAATGTTATCAGTAGAATTAAGATTTTGGCAACG 540
XX
XX 481 TTATTTATCCATTTTATCCCAATGTTATCAGTAGAATTAAGATTTTGGCAACG 540
XX 481 TTATTTATCCATTTTATCCCAATGTTATCAGTAGAATTAAGATTTTGGCAACG 540
XX
XX 541 GCACAGATTCACATTTGTTATGCCCAAAAAGATGATGATCGGTTGATGTTGCA 600
XX 541 GCACAGATTCACATTTGTTATGCCCAAAAAGATGATGATCGGTTGATGTTGCA 600
XX
XX 541 GCACAGATTCACATTTGTTATGCCCAAAAAGATGATGATCGGTTGATGTTGCA 600
XX 541 GCACAGATTCACATTTGTTATGCCCAAAAAGATGATGATCGGTTGATGTTGCA 600
XX
XX 601 GCAATTTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
XX 601 GCAATTTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
XX
XX 601 GCAATTTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
XX 601 GCAATTTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
XX
XX 661 CAGGTTCTGAAAGATGATCTGAGAGTTCCCAAGAGTTCCCAAGAGTTCCCAAGAG 720
XX 661 CAGGTTCTGAAAGATGATCTGAGAGTTCCCAAGAGTTCCCAAGAGTTCCCAAGAG 720
XX
XX 661 CAGGTTCTGAAAGATGATCTGAGAGTTCCCAAGAGTTCCCAAGAGTTCCCAAGAG 720
XX 661 CAGGTTCTGAAAGATGATCTGAGAGTTCCCAAGAGTTCCCAAGAGTTCCCAAGAG 720
XX
XX 721 AACTGGGAATTCAAACATGAAAGATGATGATGATGATGATGATGATGATGATGAT 780
XX 721 AACTGGGAATTCAAACATGAAAGATGATGATGATGATGATGATGATGATGATGAT 780
XX
XX 721 AACTGGGAATTCAAACATGAAAGATGATGATGATGATGATGATGATGATGATGAT 780
XX 721 AACTGGGAATTCAAACATGAAAGATGATGATGATGATGATGATGATGATGATGAT 780
XX
XX 781 GACGTCAAGGCTGCTCGAAGAACACCAATTTGATATCAAGATCTCCATGGAAG 840
XX 781 GACGTCAAGGCTGCTCGAAGAACACCAATTTGATATCAAGATCTCCATGGAAG 840
XX
XX 781 GACGTCAAGGCTGCTCGAAGAACACCAATTTGATATCAAGATCTCCATGGAAG 840
XX 781 GACGTCAAGGCTGCTCGAAGAACACCAATTTGATATCAAGATCTCCATGGAAG 840
XX
XX 781 GACGTCAAGGCTGCTCGAAGAACACCAATTTGATATCAAGATCTCCATGGAAG 840
XX 781 GACGTCAAGGCTGCTCGAAGAACACCAATTTGATATCAAGATCTCCATGGAAG 840
XX

```


Qy	841	GAAGAGCAGAGAAAGCTCTGTTGTGTATGACACAGTTAATAGTGCCTAATTTACAGTTT	900
Db	929	GAAGAGCCAGAGAAAGAGCTCTGTGTGTATGACACAGTTAATAGTGCCTAATTTACAGTTT	988
Qy	901	ATGAGAGAAATTGAGGGGAAATGCGTGAAAAATTCGACTACGCCAGAGACTTATATTAA	960
Db	989	ATGAGAGAAATTGAGGGGAAATGCGTGAAAAATTCGACTACGCCAGAGACTTATATTAA	1048
Qy	961	GAGTTAATCATTTCTGTGAGCCTTTGACTGTTCGATTAGACATCAGAAAAGGTTA	1022
Db	1049	GAGTTAATCATTTCTGTGAGCCTTTGACTGTTCGATTAGACATCAGAAAAGGTTA	1108
Qy	1021	CAAGCATTTAAACAAGAAATCAGAGGTCCAAATTGAACCTGATGCCAAACCCAGTTGTTG	1080
Db	1109	CAAGCATTTAAACAAGAAATCAGAGGTCCAAATTGAACCTGATGCCAAACCCAGTTGTTG	1168
Qy	1081	GACCGTTTGCAGAGAGTTCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	1140
Db	1169	GACCGTTTGCAGAGAGTTCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	1228
Qy	1141	GATGCATATGCTGTATTTAGTGTGAAAAATCAATGCGAAATTTTAAGCAGAAAACCTTT	1200
Db	1229	GATGCATATGCTGTATTTAGTGTGAAAAATCAATGCGAAATTTTAAGCAGAAAACCTTT	1288
Qy	1201	GAAGATCCAGATTAATTTTCAATATGTTTACTGCGTTGATTTGGAGAGCAACAGAAAGT	1260
Db	1289	GAAGATCCAGATTAATTTTCAATATGTTTACTGCGTTGATTTGGAGAGCAACAGAAAGT	1348
Qy	1261	GTACTTGAAGAAAAACCGAGAGACTATATAGCTTTATAA	1299
Db	1349	GTACTTGAAGAAAAACCGAGAGACTATATAGCTTTATAA	1387

CC	XX	RESULT 3
CC	XX	AAS23422
CC	ID	AAS23422 standard; DNA; 1299 BP.
CC	AC	AAS23422;
CC	DT	
CC	XX	04-DEC-2001 (first entry)
CC	DE	Candida albicans essential gene CAYMR220W (ERG8).
CC	KM	Gene identification; essential gene; GRACE; pathogenic fungus;
CC	XX	gene replacement and conditional expression; fungal infection; ds.
CC	OS	Candida albicans.
CC	PN	WO200160975-A2.
CC	XX	
CC	PD	23-AUG-2001.
CC	Pf	20-FEB-2001; 2001WO-US005551.
CC	PR	18-FEB-2000; 2000US-0183534P.
CC	XX	(ELIT-) ELITRA PHARM INC.
CC	PPA	
CC	XX	
CC	Pt	Roemer T, Jiang B, Boone C, Bussey H;
CC	DR	WPI; 2001-489080/53.
CC	XX	P-PsDB; AAU15094.
CC	PT	Identifying genes essential to fungal metabolisms and identifying
CC	XX	potential therapeutic agents that target these genes.
CC	S8	Claim 22; Page 162; 324pp; English.
CC	XX	
CC	XX	The present invention relates to novel methods for constructing fungal
CC	XX	strains useful for identification and validation of gene products as
CC	XX	targets for therapeutic agents, for creating a collection of identified
CC	XX	essential genes, and screening assays for the discovery of new drugs. The
CC	XX	invention provides the GRACE (gene replacement and conditional

Query Match	Best Local Similarity	Score	DB	Length	Mismatches	Indels	Gaps
1	ATGTCAAAAGCATTTAGTGCACCTGGAAAAGCATTTCTTCTGTGTGATATTTGGTCTT	60					
1	ATGTCAAAAGCATTTAGTGCACCTGGAAAAGCATTTCTTCTGTGTGATATTTGGTCTT	60					
1	ATGTCAAAAGCATTTAGTGCACCTGGAAAAGCATTTCTTCTGTGTGATATTTGGTCTT	60					
61	GAGCCAAATTTAGTGCACCTGGAAAAGCATTTCTTCTGTGTGATATTTGGTCTT	120					
61	GAGCCAAATTTAGTGCACCTGGAAAAGCATTTCTTCTGTGTGATATTTGGTCTT	120					
121	CCAAAAGGACAGTTTGAAGAAATTTCTTCAACCCCAATTTGCAAC	180					
121	CCAAAAGGACAGTTTGAAGAAATTTCTTCAACCCCAATTTGCAAC	180					
181	GGAGATGGGAATTCATATATCATCAATATCAAGAAAGGATTCAGTCAAGC	240					
181	GGAGATGGGAATTCATATATCATCAATATCAAGAAAGGATTCAGTCAAGC	240					
241	ATTAATTCATTTTGAAGGACATATATCATCAATATCAAGAAAGGATTCAGTCAAGC	300					
241	ATTAATTCATTTTGAAGGACATATATCATCAATATCAAGAAAGGATTCAGTCAAGC	300					
241	ATTAATTCATTTTGAAGGACATATATCATCAATATCAAGAAAGGATTCAGTCAAGC	300					
301	GCAITTTGATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	360					
301	GCAITTTGATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	360					
361	GAATTCATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	420					
361	GAATTCATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	420					
421	GAATTCATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	480					
421	GAATTCATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	480					
481	TTATTCATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	540					
481	TTATTCATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	540					
541	GCACAGATTCGATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	600					
541	GCACAGATTCGATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	600					
601	GCAITTTGATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	660					
601	GCAITTTGATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	660					
661	CAGGTTCTGAAAGTATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	720					
661	CAGGTTCTGAAAGTATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	720					
721	AACTGGAAATTCGAAAGTATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	780					
721	AACTGGAAATTCGAAAGTATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	780					
781	GAGGTTCTGAAAGTATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	840					
781	GAGGTTCTGAAAGTATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	840					
841	GAAGGTTCTGAAAGTATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	900					
841	GAAGGTTCTGAAAGTATCTTGAATCATCATTTACTGACACCTGATATCATTCACAAGAAAGTACT	900					

QY 901 ATGAGGAATGAGGAGAAATGCGTGAATAATACGACTCAGACCCAGACTTATATTAA 960
DB 901 ATGAGGAATGAGGAGAAATGCGTGAATAATACGACTCAGACCCAGACTTATATTAA 960
QY 961 GAGTTAGATCATTTCTGTTGAGCCTTTGACTGTTGCGATTAAACATCAGAAAAGGTTA 1020
DB 961 GAGTTAGATCATTTCTGTTGAGCCTTTGACTGTTGCGATTAAACATCAGAAAAGGTTA 1020
QY 1021 CAAGCATTAAACAAAAATCAGAGGTTCAATTGAACTGATGTCGCAAAACCAAGTTGTG 1080
DB 1021 CAAGCATTAAACAAAAATCAGAGGTTCAATTGAACTGATGTCGCAAAACCAAGTTGTG 1080
QY 1081 GACCGTTGTCAGAGATTCCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1140
DB 1081 GACCGTTGTCAGAGATTCCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1140
QY 1141 GATGCAATAGCTGATTAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1200
DB 1141 GATGCAATAGCTGATTAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1200
QY 1201 GAAATATCAGATTATTTTCAATTAATGTTTACTGGGTTGATTTTGAAGAGCAAAAGT 1260
DB 1201 GAAATATCAGATTATTTTCAATTAATGTTTACTGGGTTGATTTTGAAGAGCAAAAGT 1260
QY 1261 GTACTGAGAGAAAACCAAGAACTATATAGTTTATTA 1299
DB 1261 GTACTGAGAGAAAACCAAGAACTATATAGTTTATTA 1299

RESULT 4
ID AB231736 standard; DNA; 1299 BP.
XX AB231736;

DT 30-JAN-2003 (first entry)
XX
DE Candida albicans essential gene SEQ ID NO 6023.
XX
KM Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
KM signal transduction; DNA replication; cell division; growth;
KM proliferation; Candida albicans; fungicide; antifungal; gene; ss.
XX
OS Candida albicans.
XX
PN WO200253728-A2.
XX
PD 11-JUL-2002.
XX
PF 26-DEC-2001; 2001WO-US049486.
XX
PR 29-DEC-2000; 2000US-0259128P.
PR 20-FEB-2001; 2001US-00792024.
PR 22-AUG-2001; 2001US-0314050P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL,
XX
XX WPI; 2002-566694/60.
DR P-PSDB; ABP73186.
XX
PT Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele of
PT a gene and placing other allele of the gene under conditional expression.
XX
XX Claim 37; SEQ ID NO 6023; 167bp + Sequence Listing; English.
XX
XX The invention relates to constructing (M1) a strain of diploid fungal
XX cells in which both alleles of a gene are modified, comprising modifying
XX one allele by insertion or replacement by a cassette having an
XX expressible selectable marker and modifying other allele by
XX recombination, of a promoter replacement fragment with a heterologous

CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication, and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans gene used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office
XX
SQ Sequence 1299 BP; 439 A; 221 C; 271 G; 368 T; 0 U; 0 Other;
Query Match 68.6%; Score 891; DB 6; Length 1299;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1291; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 ATGTCAAAAGCATTATAGTGCACCTGGAAAAGCATTTCTGCTGCTGATATTTGGTTCTT 60
DB 1 ATGTCAAAAGCATTATAGTGCACCTGGAAAAGCATTTCTGCTGCTGATATTTGGTTCTT 60
QY 61 GAGCAATTTATAGATGCTTATGTCAGCATTTGTCATCAGAAATGATGAGTTATTA 120
DB 61 GAGCAATTTATAGATGCTTATGTCAGCATTTGTCATCAGAAATGATGAGTTATTA 120
QY 121 CCAAAAGAACCAATTTGAAAGATCTAGATCAAAATTTCTTACCCCAATTTGCAAC 180
DB 121 CCAAAAGAACCAATTTGAAAGATCTAGATCAAAATTTCTTACCCCAATTTGCAAC 180
QY 181 GAGCATGGAATATATCATATCATATCAATATCAGAGAGCCGAGAGATTCAGTCAGC 240
DB 181 GAGCATGGAATATATCATATCATATCAATATCAGAGAGCCGAGAGATTCAGTCAGC 240
QY 181 GAGGAATGGGAATATCATATCATATCAATATCAGAGAGCCGAGAGATTCAGTCAGC 240
DB 181 GAGGAATGGGAATATCATATCATATCAATATCAGAGAGCCGAGAGATTCAGTCAGC 240
QY 241 ATAAATCATTTTATAGAGCACTATATATGCTTATATATCAACGAGCGAA 300
DB 241 ATAAATCATTTTATAGAGCACTATATATGCTTATATATCAACGAGCGAA 300
QY 301 GCATTTGATCTTGAATATCATATTTACTGAGACCTGATATGATTCACAAGAGATAT 360
DB 301 GCATTTGATCTTGAATATCATATTTACTGAGACCTGATATGATTCACAAGAGATAT 360
QY 361 GAAACCAAGATCTCTGATGAGAGAAAACATTTCTTATCAATTTCTGTCATTAAC 420
DB 361 GAAACCAAGATCTCTGATGAGAGAAAACATTTCTTATCAATTTCTGTCATTAAC 420
QY 421 GAAATGGAAGAACCGGATAGGTTATGTCAGAGATTTGTCAGTTGTCACAAGT 480
DB 421 GAAATGGAAGAACCGGATAGGTTATGTCAGAGATTTGTCAGTTGTCACAAGT 480
QY 481 TTATTAATCCATTTTATCCCAATGTTATCAGATGATTAAGATATTTTGCACAAAGT 540
DB 481 TTATTAATCCATTTTATCCCAATGTTATCAGATGATTAAGATATTTTGCACAAAGT 540
QY 541 GCAAGATGTCAGATGTTATGTCGCAAAAAGATGATGCTGGTTGATGTTGCAACT 600
DB 541 GCAAGATGTCAGATGTTATGTCGCAAAAAGATGATGCTGGTTGATGTTGCAACT 600
QY 601 GCAATTTATGCTGATGTTATGATTAAGATTTTCAAGCTTGAATTAATGACGTTT 660
DB 601 GCAATTTATGCTGATGTTATGATTAAGATTTTCAAGCTTGAATTAATGACGTTT 660
QY 661 CAGGTTCTAGAAAGTATCTGGAAGTTCCCAAGAGTTGAAAAATTTGAAAGT 720
DB 661 CAGGTTCTAGAAAGTATCTGGAAGTTCCCAAGAGTTGAAAAATTTGAAAGT 720


```

Db      661 CAGGTTCTAGAAAGTATCTGAGAGTTCCCAAGATGTAATAATTGATGCAAGT 720
Qy      721 AACTGGGAATTCACACATGAAAGATGTAACCTTACCAACGGAATCAAGTATTAATGAGT 780
Db      721 AACTGGGAATTCACACATGAAAGATGTAACCTTACCAACGGAATCAAGTATTAATGAGT 780
Qy      781 GAGCTCAAGGGTGTGCTAGAAACACCCAAATTTGTTACGAGTACTCCAAATGGAAGAG 840
Db      781 GAGCTCAAGGGTGTGCTAGAAACACCCAAATTTGTTACGAGTACTCCAAATGGAAGAG 840
Qy      841 GAAAGCCGACAAAGAAAGCTTGTGTGTATGACAGCTTAAATAGTCCAAATTTAAGATT 900
Db      841 GAAAGCCGACAAAGAAAGCTTGTGTGTATGACAGCTTAAATAGTCCAAATTTAAGATT 900
Qy      901 ATGAGGAATTGAGGAAATGCGTGAATAATGAGCTGAGCCGAGACTTATATTTAA 960
Db      901 ATGAGGAATTGAGGAAATGCGTGAATAATGAGCTGAGCCGAGACTTATATTTAA 960
Qy      961 GAGTTGATCATTTCTGTTGAGCCCTTTGACTGTGCGATTAGACATCAGAAAAGGTTA 1020
Db      961 GAGTTGATCATTTCTGTTGAGCCCTTTGACTGTGCGATTAGACATCAGAAAAGGTTA 1020
Qy      1021 CAAGCATTAACAACAAATATCAGAGTCCAAATGAACTGATGTCCAAACCAAGTTGTG 1080
Db      1021 CAAGCATTAACAACAAATATCAGAGTCCAAATGAACTGATGTCCAAACCAAGTTGTG 1080
Qy      1081 GACCGTTGTCAAAGATTCCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
Db      1081 GACCGTTGTCAAAGATTCCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
Qy      1141 GATGCAATAGCTGTATATGATGTGTGAAATCAAGTGGGAAATTTTAAAGCAAAAACTCT 1200
Db      1141 GATGCAATAGCTGTATATGATGTGTGAAATCAAGTGGGAAATTTTAAAGCAAAAACTCT 1200
Qy      1201 GAAAAATCCAGATTATTTTCATATGTTACTGGGTGATTTGGAAGCAAAACGAAAGT 1260
Db      1201 GAAAAATCCAGATTATTTTCATATGTTACTGGGTGATTTGGAAGCAAAACGAAAGT 1260
Qy      1261 GTACTTGAAGAAAAACGAGAACTATATAGTTTATPA 1299
Db      1261 GTACTTGAAGAAAAACGAGAACTATATAGTTTATPA 1299

RESULT 5
AAd02787
ID      AAd02787 standard; DNA; 547 BP.
AC      AAd02787;
DT      31-MAY-2001 (first entry)
XX      S. cerevisiae ERG8 gene homologous DNA #1 from Candida albicans.
XX      Phosphomevalonate kinase; PMK; ERG8; anti-fungal agent; diagnosis;
XX      infection; ds.
XX      Candida albicans.
XX      WO200114533-A2.
XX      01-MAR-2001.
XX      15-AUG-2000; 2000WO-GB003100.
XX      21-AUG-1999; 99GB-00019766.
XX      (ASTR ) ASTRAZENECA AB.
XX      (ASTR ) ASTRAZENECA UK LTD.
XX      Rosamond JDC, Schnell NF;
XX      WPI; 2001-218441/22.
XX

```

```

PT      New polypeptides and polynucleotides (ERG8) from Candida albicans, useful
PT      in assays for identifying inhibitors of phosphomevalonate kinase activity
PT      and as reagents for diagnosing C. albicans infection.
XX      Claim 6; Page 24; 29pp; English.
XX
XX      The patent discloses phosphomevalonate kinase (PMK; ERG8) protein and
XX      their corresponding DNAs from Candida albicans. The ERG8 protein is
XX      useful in assays for identifying compounds that inhibit phospho-
XX      mevalonate kinase (PMK) activity. These inhibitors are useful as anti-
XX      fungal agents. The ERG8 DNA and protein are also useful as reagents for
XX      diagnosing C. albicans infection. The present sequence is a Candida
XX      albicans DNA which is homologous to the Saccharomyces cerevisiae ERG8
XX      gene
SQ      Sequence 547 BP; 184 A; 81 C; 123 G; 159 T; 0 U; 0 Other;
Qy      828 CCATGGAAGAAAGCAAGAAAGCTCTGTGTATGACCACTTAATATGTC 887
Db      1 CCAATGGAAGAAAGCAAGAAAGCTCTGTGTATGACCACTTAATATGTC 60
Qy      888 CAATTTACGTTTATGAGAAATGAGGAAATGCGTGAATAATGCACTCAGACCCAGA 947
Db      61 CAATTTACGTTTATGAGAAATGAGGAAATGCGTGAATAATGCACTCAGACCCAGA 120
Qy      948 GACTTATATTAAGAGTTAGATCAATCTGTGAGCTTTGACTGTGCGATTAAAGACAT 1007
Db      121 GACTTATATTAAGAGTTAGATCAATCTGTGAGCTTTGACTGTGCGATTAAAGACAT 180
Qy      1008 CAGAAAGGGTTACCAAGATTACACAAAATCAGAGTTCAATTGAACCTGATGCCA 1067
Db      181 CAGAAAGGGTTACCAAGATTACACAAAATCAGAGTTCAATTGAACCTGATGCCA 240
Qy      1068 AACCGATGTTGACCGCTTGTCAAGAGATTCCTGTTGTGTGTGTGTGTGTGTGTGTGT 1127
Db      241 AACCGATGTTGACCGCTTGTCAAGAGATTCCTGTTGTGTGTGTGTGTGTGTGTGTGT 300
Qy      1128 TGTGTGTGATACGATGCAATAGCTGTATTAAGTGTGAAATCAAGTGGGAAATTTTAA 1187
Db      301 TGTGTGTGATACGATGCAATAGCTGTATTAAGTGTGAAATCAAGTGGGAAATTTTAA 360
Qy      1188 GCAGAAATCTTTGAAATTCAGATTATTTTCATATGTTACTGGGTGATTTGAAAGA 1247
Db      361 GCAGAAATCTTTGAAATTCAGATTATTTTCATATGTTACTGGGTGATTTGAAAGA 420
Qy      1248 GCAAAACGAAGTGTACTTGAAGAAACGAGAACTATATAGTTTATPA 1299
Db      421 GCAAAACGAAGTGTACTTGAAGAAACGAGAACTATATAGTTTATPA 472

RESULT 6
AAd02788/c
ID      AAd02788 standard; DNA; 577 BP.
AC      AAd02788;
DT      31-MAY-2001 (first entry)
XX      S. cerevisiae ERG8 gene homologous DNA #2 from Candida albicans.
XX      Phosphomevalonate kinase; PMK; ERG8; anti-fungal agent; diagnosis;
XX      infection; ds.
XX      Candida albicans.
XX      WO200114533-A2.
XX      01-MAR-2001.
XX

```


XX Tetracycline promoter downstream PCR primer (Tet-Down) #42.
 DE
 XX
 KW Gene identification; essential gene; GRACE; pathogenic fungus;
 KW gene replacement and conditional expression; fungal infection;
 KW PCR primer; Tet-Down; tetracycline promoter; ss.
 XX
 OS Escherichia coli.
 OS Candida albicans.
 XX
 PN WO200160975-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US005551.
 XX
 PR 18-FEB-2000; 2000US-0183534P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Roemer T, Jiang B, Boone C, Buesey H;
 XX
 DR WPI; 2001-489080/53.
 XX
 PT Identifying genes essential to fungal metabolisms and identifying
 PT potential therapeutic agents that target these genes.
 XX
 PS Disclosure; Page 297; 324pp; English.
 XX
 CC The present invention relates to novel methods for constructing fungal
 CC strains useful for identification and validation of gene products as
 CC targets for therapeutic agents, for creating a collection of identified
 CC essential genes, and screening assays for the discovery of new drugs. The
 CC invention provides the GRACE (gene replacement and conditional
 CC expression) method for the construction of mutant organisms referred to
 CC as GRACE strains of the organism. The invention can be applied to any
 CC organism, particularly a pathogenic fungus e.g. Candida albicans,
 CC Aspergillus fumigatus and Cryptococcus neoformans. The methods are useful
 CC to identify agents that may be used in the treatment of fungal
 CC infections. AAS23626-AAS23686 represent tetracycline promoter downstream
 CC PCR primers (Tet-Down) used in the methods of the present invention
 XX
 SQ Sequence 90 BP; 28 A; 19 C; 18 G; 25 T; 0 U; 0 Other;
 Query Match 5.0%; Score 65; DB 4; Length 90;
 Best Local Similarity 100.0%; Pred. No. 5.9e-21;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGTCAAAAGCATTAGTGCACCTGAAAGCATTCTTGCTGCGATATTTGGTTCTT 60
 DB 65 ATGTCAAAAGCATTAGTGCACCTGAAAGCATTCTTGCTGCGATATTTGGTTCTT 6
 QY 61 GAGCC 65
 DB 5 GAGCC 1
 XX
 PN WO200114533-A2.
 XX
 DE
 XX
 KW Phosphomevalonate kinase; PMK; ERG8; anti-fungal agent; diagnosis;
 KW infection; PCR primer; ss.
 XX
 OS Candida albicans.
 XX
 PN WO200114533-A2.

XX
 PD 01-MAR-2001.
 XX
 PF 15-AUG-2000; 2000WO-GB003100.
 XX
 PR 21-AUG-1999; 99GB-00019766.
 XX
 PA (ASTR) ASTRAZENECA AB.
 PA (ASTR) ASTRAZENECA UK LTD.
 XX
 PI Rosamond JDC, Schnell NF;
 XX
 DR WPI; 2001-218441/22.
 XX
 PT New polypeptides and polynucleotides (ERG8) from Candida albicans, useful
 PT in assays for identifying inhibitors of phosphomevalonate kinase activity
 PT and as reagents for diagnosing C. albicans infection.
 XX
 PS Example 1; Page 25; 29pp; English.
 XX
 CC The patent discloses phosphomevalonate kinase (PMK; ERG8) protein and
 CC their corresponding DNAs from Candida albicans. The ERG8 protein is
 CC useful in assays for identifying compounds that inhibit phospho-
 CC mevalonate kinase (PMK) activity. These inhibitors are useful as anti-
 CC fungal agents. The ERG8 DNA and protein are also useful as reagents for
 CC diagnosing C. albicans infection. The present DNA sequence is a PCR
 CC primer which is used in the isolation of a clone containing the Candida
 CC albicans ERG8 gene
 XX
 SQ Sequence 25 BP; 11 A; 1 C; 7 G; 6 T; 0 U; 0 Other;
 Query Match 1.9%; Score 25; DB 4; Length 25;
 Best Local Similarity 100.0%; Pred. No. 0.18;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 477 AAGTTATTATTCACATTATTCGCC 501
 DB 25 AAGTTATTATTCACATTATTCGCC 1
 XX
 PN WO200114533-A2.
 XX
 PD 01-MAR-2001.
 XX
 PF 15-AUG-2000; 2000WO-GB003100.
 XX
 PR 21-AUG-1999; 99GB-00019766.
 XX
 PA (ASTR) ASTRAZENECA AB.
 PA (ASTR) ASTRAZENECA UK LTD.
 XX
 PI Rosamond JDC, Schnell NF;
 XX
 DR WPI; 2001-218441/22.
 XX
 PT New polypeptides and polynucleotides (ERG8) from Candida albicans, useful
 PT in assays for identifying inhibitors of phosphomevalonate kinase activity
 PT and as reagents for diagnosing C. albicans infection.

XX Example 4; Page 29; 29pp; English.
PS
CC The patent discloses phosphomevalonate kinase (PMK; ERG8) protein and
CC their corresponding DNAs from *Candida albicans*. The ERG8 protein is
CC useful in assays for identifying compounds that inhibit phospho-
CC mevalonate kinase (PMK) activity. These inhibitors are useful as anti-
CC fungal agents. The ERG8 DNA and protein are also useful as reagents for
CC diagnosing *C. albicans* infection. The present DNA sequence is PCR primer
CC which is used to amplify the *Candida albicans* ERG8 coding sequence. This
CC sequence incorporates restriction enzyme sites in the ERG8 coding
CC sequence and facilitate its cloning
XX
SQ Sequence 36 BP; 9 A; 8 C; 6 G; 13 T; 0 U; 0 Other;

Query Match 1.8%; Score 24; DB 4; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1276 CCAGAAAGCTATATAGGTTTATAA 1299
Db |||||
36 CCAGAAAGCTATATAGGTTTATAA 13

RESULT 11
AAS23789/c
ID AAS23789 standard; DNA; 22 BP.
XX
AC AAS23789;
XX
DT 04-DEC-2001 (first entry)
XX
DE Primer B #42 used as probe for identifying *C. albicans* GRACE strain.
XX
KW Gene identification; essential gene; GRACE; pathogenic fungus;
KW gene replacement and conditional expression; fungal infection; probe; ss.
XX
OS *Candida albicans*.
OS Synthetic.
XX
PN WO200160975-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005551.
XX
PR 18-FEB-2000; 2000US-0183534P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H;
XX
DR WPI; 2001-489080/53.
XX
PT Identifying genes essential to fungal metabolisms and identifying
PT potential therapeutic agents that target these genes.
XX
PS Claim 36; Page 320; 324pp; English.
XX
CC The present invention relates to novel methods for constructing fungal
CC strains useful for identification and validation of gene products as
CC targets for therapeutic agents, for creating a collection of identified
CC essential genes, and screening assays for the discovery of new drugs. The
CC invention provides the GRACB (gene replacement and conditional
CC expression) method for the construction of mutant organisms referred to as
CC as GRACB strains of the organism. The invention can be applied to any
CC organism, particularly a pathogenic fungus e.g. *Candida albicans*,
CC *Aspergillus fumigatus* and *Cryptococcus neoformans*. The methods are useful
CC to identify agents that may be used in the treatment of fungal
CC infections. AAS23748-AAS23808 represent primers B #1-61 used as probes
CC for identifying *C. albicans* GRACB strains
XX
SQ Sequence 22 BP; 8 A; 7 C; 3 G; 4 T; 0 U; 0 Other;

Query Match 1.7%; Score 22; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 GGATTAGTGTGACGTTGTGCCA 475
Db |||||
22 GGATTAGTGTGACGTTGTGCCA 1

RESULT 12
ABZ30804/c
ID ABZ30804 standard; DNA; 22 BP.
XX
AC ABZ30804;
XX
DT 30-JAN-2003 (first entry)
XX
DE *Candida albicans* GRACE strain PCR primer SEQ ID NO 5023.
XX
KW Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth; ss.
KW proliferation; *Candida albicans*; fungicide; antifungal; PCR; primer; ss.
XX
OS *Candida albicans*.
OS
XX
PN WO200253728-A2.
XX
PD 11-JUL-2002.
XX
PF 26-DEC-2001; 2001WO-US049486.
XX
PR 29-DEC-2000; 2000US-0259128P.
PR 20-FEB-2001; 2001US-00792024.
PR 22-AUG-2001; 2001US-0314050P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
XX
DR WPI; 2002-566694/60.
XX
PT Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele of
PT a gene and placing other allele of the gene under conditional expression.
XX
PS Claim 36; SEQ ID NO 5023; 167pp + Sequence Listing; English.
XX
CC The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a diploid fungal
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of *C. albicans* cells and for
CC treating infection by *C. albicans*. The present sequence is that of a PCR
CC primer used in the method of the invention. Note: The sequence data for
CC this patent is not represented in the printed specification but is based
CC on sequence information supplied to Derwent by the European Patent Office
XX

AC ACH30537

Claim 1; SEQ ID NO 17749; 44pp; English

The invention relates to an isolated polynucleotide comprising any one of 38043 cDNA sequences, appearing as ACHT2789-ACH50833, whose sequence was determined by the technique of SSH (sequencing by hybridisation). Also included is a purified polypeptide comprising a sequence corresponding to a reading frame of the novel polynucleotide. The nucleic acid sequences are useful in diagnostics as expressed sequence tags (EST) for identifying expressed genes or for physical mapping of the human genome, in forensics, in assessing biodiversity, or in identifying mutations responsible for genetic disorders and other traits. The nucleotide sequences are also useful as hybridisation probes, as oligomers for PCR, for chromosome and gene mapping, in the recombinant production of protein, or in generating antisense DNA or RNA. The purified polypeptide is useful for generating antibodies specific for it. The present sequence is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030073623

Query Match	1.68;	Score 21;	DB 9;	Length 476;
Best Local Similarity	100.0%	21	9	476

Matches	21;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0
---------	-----	--------------	----	------------	----	--------	----	------	---

1115 GTGTGTTCCAGGTGCTGGTG 1135

Db 126 GTGTGTTCCAGGTGCTGGTG 146

RESULT 15
ABK45175

APR 7 1964

05-JUN-2002 (first entry)

CDNA encoding colon tumour protein, SEQ ID No 726.

KM Human; colon tumour; vaccine; colon cancer; immunogenic; immunotherapy;
KM gene; ss.
XX
XX Homo sapiens.
OS
XX WO200212328-A2.
PN
XX
XX 14-FEB-2002.
PD
XX
XX 31-JUL-2001; 2001WO-US024218.
PF
XX
XX 03-AUG-2000; 2000US-0223283P.
PR 28-MAR-2001; 2001US-0279763P.
PR 29-JUN-2001; 2001US-0302051P.
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX King GE, Meagher MJ, Xu J, Secrist H;
PI
XX WPI; 2002-241739/29.
DR
XX
XX New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT for diagnosing, preventing, and treating colon cancer, and as markers for
PT the progression of cancer.
XX
XX Claim 1; SEQ ID NO 726; 147pp; English.
PS
XX
XX The invention relates to polynucleotides encoding colon tumour proteins.
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis, prevention, and
CC treatment of colon cancer. Polynucleotide sequences may be used as
CC hybridisation probes or primers, and in the design and preparation of
CC ribozyme molecules for inhibiting expression of tumour polypeptides and
CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer. ABK4450-
CC ABK46237 represent coding sequences of human colon tumour proteins of the
CC invention. Note: With the exception of SEQ ID No 1 and 2, the sequence
CC data for this patent did not form part of the printed specification but
CC was supplied by the European Patent Office
XX
XX Sequence 479 BP; 127 A; 93 C; 129 G; 130 T; 0 U; 0 Other;
SQ
XX
XX Query Match 1.6%; Score 21; DB 6; Length 479;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1115 GTGTGTTCCAGGTGCTGTG 1135
DB 41 GTGTGTTCCAGGTGCTGTG 61
XX
XX RESULT 16
ABV86952/c
ID ABV86952 standard; cDNA; 518 BP.
XX
XX ABEV86952;
AC
XX
XX 13-DEC-2002 (first entry)
DT
XX
XX Human colon cancer related cDNA SEQ ID NO 263.
DE
XX
XX Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
KM ss.
XX
XX Homo sapiens.
OS
XX WO200258534-A2.
PN
XX
XX 01-AUG-2002.
PD
XX
XX 16-NOV-2001; 2001WO-US043704.
PF
XX
XX

PR 20-NOV-2000; 2000US-0252222P.
PR 06-FEB-2001; 2001US-0267011P.
PR 28-MAR-2001; 2001US-0279670P.
PR 10-JUL-2001; 2001US-0304037P.
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Stolk JA, Chenault RA, Meagher MJ, Secrist H, King GE;
PI
XX WPI; 2002-608400/65.
DR
XX
XX New isolated tumor colon polynucleotide and polypeptide, useful for the
PT diagnosis, prevention and/or treatment of cancer, in particular colon
PT cancer.
XX
XX Claim 1; SEQ ID NO 263; 266pp + Sequence Listing; English.
PS
XX
XX The invention relates to a human colon tumour expressed polynucleotide
CC (i) encoding a polypeptide (II), ABP67991-ABP67996) comprising: (i) any of
CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
CC sequences that hybridize to (i), under moderately stringent conditions;
CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
CC degenerate variants of (i). The compositions, prevention and/or treatment of
CC invention are useful for the diagnosis, prevention and/or treatment of
CC cancer, particularly colon cancer. (i) can be used in gene therapy and
CC (i) and (ii) are useful in pharmaceutical compositions such as vaccines.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 518 BP; 135 A; 133 C; 98 G; 145 T; 0 U; 7 Other;
SQ
XX
XX Query Match 1.6%; Score 21; DB 6; Length 518;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1115 GTGTGTTCCAGGTGCTGTG 1135
DB 414 GTGTGTTCCAGGTGCTGTG 394
XX
XX RESULT 17
ABK45703/c
ID ABK45703 standard; cDNA; 530 BP.
XX
XX ABEK45703;
AC
XX
XX 05-JUN-2002 (first entry)
DT
XX
XX cDNA encoding colon tumour protein, SEQ ID No 1254.
DE
XX
XX Human; colon tumour; vaccine; colon cancer; immunogenic; immunotherapy;
KM gene; ss.
XX
XX Homo sapiens.
OS
XX WO200212328-A2.
PN
XX
XX 14-FEB-2002.
PD
XX
XX 31-JUL-2001; 2001WO-US024218.
PF
XX
XX 03-AUG-2000; 2000US-0223283P.
PR 28-MAR-2001; 2001US-0279763P.
PR 29-JUN-2001; 2001US-0302051P.
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX King GE, Meagher MJ, Xu J, Secrist H;
PI
XX WPI; 2002-241739/29.
DR
XX
XX

PT New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT for diagnosing, preventing, and treating colon cancer, and as markers for
XX the progression of cancer.
PS Claim 1; SEQ ID NO 1254; 147pp; English.
XX
CC The invention relates to polynucleotides encoding colon tumour proteins.
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis, prevention, and
CC treatment of colon cancer. Polynucleotide sequences may be used as
CC hybridisation probes or primers, and in the design and preparation of
CC ribozyme molecules for inhibiting expression of tumour polypeptides and
CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer. ABK4450-
CC ABK46237 represent coding sequences of human colon tumour proteins of the
CC invention. Note: With the exception of SEQ ID NO 1 and 2, the sequence
CC data for this patent did not form part of the printed specification but
CC was supplied by the European Patent Office
SQ Sequence 530 BP; 137 A; 139 C; 105 G; 149 T; 0 U; 0 Other;
Query Match 1.6%; Score 21; DB 6; Length 530;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1115 GTGGGTTCCAGGCTGCTGGTG 1135
DB 439 GTGGGTTCCAGGCTGCTGGTG 419
RESULT 18
AA197097
ID AA197097 standard; cDNA; 771 BP.
XX
AC AA197097;
XX
DT 13-NOV-2001 (first entry)
XX
DE Human neuroblastoma expressed polynucleotide SEQ ID NO 3172.
XX
KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS Homo sapiens.
XX
PN WO200166719-A1.
XX
PD 13-SEP-2001.
XX
PF 02-MAR-2001; 2001WO-JP001629.
XX
PR 07-MAR-2000; 2000JP-00159195.
XX
PA (CHIB-) CHIBA PREFECTURE.
XX (HISM) HISAMITSU PHARM CO LTD.
XX
PI Nakagawara A;
XX
PI WPI; 2001-565584/63.
XX
PT Nucleic acids originating in gene expressed in human neuroblastoma,
PT useful as probe or primer in diagnosing prognosis of human neuroblastoma,
PT malignancy and susceptibility indicator or tumor marker for anti-cancer
PT agents.
PS Claim 1; Page 2308; 2979pp; Japanese.
XX
CC The invention relates to novel genes (AA19326-AA197963) expressed in
CC human neuroblastoma. The nucleic acids are applicable as a probe or
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC susceptibility indicators or tumour markers for anti-cancer agents. The
CC gene information for diagnosing prognosis is related to factors similar
CC to that for N-myc and TrkA genes

XX
SQ Sequence 771 BP; 211 A; 143 C; 195 G; 207 T; 0 U; 15 Other;
Query Match 1.6%; Score 21; DB 4; Length 771;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1115 GTGGGTTCCAGGCTGCTGGTG 1135
DB 436 GTGGGTTCCAGGCTGCTGGTG 456
RESULT 19
AB054714
ID AB054714 standard; cDNA; 1574 BP.
XX
AC AB054714;
XX
DT 22-AUG-2002 (first entry)
XX
DE Human ovarian antigen HMSOH12 cDNA, SEQ ID NO:594.
XX
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder; infection;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2002006771-A1.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WO-US018569.
XX
PR 07-JUN-2000; 2000US-0209467P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
PI WPI; 2002-147878/19.
XX P-ESDB; ABP41637.
XX
DR
XX
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
PT cancer), immune disorders, cardiovascular disorders and neurological
PT diseases.
PS Claim 1; SEQ ID NO 594; 2922pp; English.
XX
CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (AB054131-AB056105), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovarian cancer and breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),

CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents cDNA encoding a human ovarian antigen of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp:wipo.int/pub/published_pat_sequences

Sequence 1574 BP; 501 A; 251 C; 359 G; 455 T; 0 U; 8 Other;

Query Match	1.6%	Score 21	DB 6	Length 1574
Best Local Similarity	100.0%	Pred. No. 13		
Matches 21; Conservative	0	Mismatches	0	Gaps 0

```

QY      1115 GTGTGTTCCAGGTGCTGTG 1135
          |||||
Db      701 GTGTGTTCCAGGTGCTGTG 721

```

RESULT 20

ID	ADQ22869	standard; DNA; 1825 BP.
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

AC ADQ22869;

DT 26-AUG-2004 (First entry)

DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5689.

KM soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KM ds.

OS Homo sapiens.

PN WO2004048938-A2.

PD 10-JUN-2004

PF 26-NOV-2003; 2003WO-US038193.

PR 26-NOV-2002; 2002US-0429739P.

PA (PROT-) PROTEIN DESIGN LABS INC.

PI Aziz N, Ginsburg WM, Zlotnick A;

DR WPI; 2004-441208/41.

PT Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample PT and comparing the gene expression, also useful in treating soft tissue PT sarcoma.

PS Example 2; SEQ ID NO 5689; 210pp; English.

The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytotoxic applications and may be useful for detecting soft tissue sarcoma, possibly via gene therapy or vaccine production. The nucleic acid sequences may be useful in diagnostic and screening applications. The current sequence is that of a human soft tissue sarcoma-upregulated DNA of the invention. The current sequence is not shown within the

CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence, 1825 BP; 500 A; 379 C; 415 G; 531 T; 0 U; 0 Other;

Sequence 1825 BP; 500 A; 379 C; 415 G; 531 T; 0 U; 0 Other;

Query Match Similarity 1.6%; Score 21; DB 12; Length 1625;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 115 GTGTGTTCCAGGTGCTGGTG 1135
|||||
|||
Db 448 GTGTGTTCCAGGTGCTGGTG 468

RESULT 21

ID ABV78079 standard; DNA; 2010 BP.

AC ABV78079 ;

DT 12-NOV-2002 (first entry)

Hypoxia-regulated protein coding sequence #99.

KM Cytostatic; vasotrophic; tranquiliser; antiatherosclerotic; gene therapy;
KM antiinflammatory; vulnerezy; gynecological; ophthalmological; vaccine;
KM hypoxia; tumorigenesis; angiogenesis; apoptosis; cancer;
KM ischemic condition; reperfusion injury; retinopathy; neonatal stress;
KM preeclampsia; atherosclerosis; inflammatory condition; wound healing;
KM inflammation; erythropoiesis; hair loss; human; gene; ds.

OS Homo sapiens.

PN WO200246465-A2.

PD 13-JUN-2002

PF 10-DEC-2001; 2001WO-GB005458.

PR 08-DEC-2000; 2000GB-00030076.

PR 25-OCT-2001; 2001GB-00025666.

PA (OXFO-) OXFORD BIOMEDICA UK I

PI White J, Mundy CR, Ward NR, K

XX
XX
XXXXXXX/

XXXX
XXXX
XXXX
XXXX
XXXX

PT conditions, comprised

PT regulated gene:

PS Claim 37; Page 403-404; 538pp; English.

The present invention relates to methods for identifying genes and proteins that are implicated in a specific disease or physiological condition. The method comprises comparing the transcriptome/proteome of a specialised cell type implicated in a disease or condition with that of a second specialised cell type, under two experimental conditions, and identifying a gene that is differentially regulated in the two specialised cell types under experimental conditions. ABV77813-ABV78116 and ABP5061-ABP6257 were identified using the methods of the invention. The coding sequences and proteins are useful for treating a disease in a patient, for manufacture of a medicament for treating hypoxia-regulated conditions, and for regulating tumorigenesis, angiogenesis, apoptosis, biological response to hypoxia conditions, or hypoxia-associated pathology in a patient. The coding sequences and proteins are also useful for monitoring the therapeutic treatment of a disease or physiological condition, such as cancer, ischaemic conditions, reperfusion injury, retinopathy, neonatal stress, preeclampsia, atherosclerosis, inflammatory conditions, wound healing, inflammation, erythropoiesis or hair loss

XX SQ Sequence 2010 BP; 607 A; 388 C; 498 G; 517 T; 0 U; 0 Other;
 CC Query Match 1.6%; Score 21; DB 6; Length 2010;
 CC Best Local Similarity 100.0%; Pred. No. 13;
 CC Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1115 GTGTGTTCCAGTGTCTGTG 1135
 DB 1303 GTGTGTTCCAGTGTCTGTG 1323

RESULT 22
 ADCC29939
 ID ADCC29939 standard; cDNA; 2461 BP.
 XX AC ADCC29939;
 XX XX
 DT 18-DEC-2003 (first entry)
 XX DE Human novel cDNA sequence, SEQ ID NO:21.
 XX KW Human, diagnostic; drug screening; forensics; gene mapping;
 KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
 KW neurodegenerative diseases; anemia; platelet disorder; wound; burns;
 KW ulcers; osteoporosis; autoimmune disease; cancer;
 KW molecular weight marker; food supplement; antiparkinsonian; nocotropic;
 KW neuroprotective; anti-nausea; anticoagulant; thrombolytic; vulnery;
 KW anticancer; osteopathic; immunosuppressive; antiinflammatory; cytosolitic;
 KW gene therapy; gene; ss.
 XX OS Homo sapiens.
 XX PN WO2003029271-A2.
 XX PD 10-APR-2003.
 XX XX
 PF 24-SEP-2002; 2002WO-US030474.
 XX PR 24-SEP-2001; 2001US-032463P.
 XX PA (HYGE-) HYSEQ INC.
 XX PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
 PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
 PI Haley-Vicente D, Drmanac RT;
 XX WPI; 2003-371981/35.
 DR P-PSDB; ADCC30910.
 XX DE
 XX PT New polynucleotide and polypeptide useful for diagnosing, preventing or
 PT treating conditions such as neurodegenerative diseases, anemias, platelet
 PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 PT cancer.
 XX PT
 XX Claim 1, SEQ ID NO 21, 1185bp; English.
 XX PS
 XX XX
 CC The invention relates to 971 novel human cDNA sequences (ADCC29919-
 CC ADCC30889) and the polypeptides they encode (ADCC30890-ADCC31860). The
 CC invention also relates to nucleic acid sequences over 99% identical with
 CC the novel human cDNAs. The invention additionally encompasses expression
 CC vectors and host cells comprising a nucleic acid of the invention; the
 CC recombinant production of a polypeptide of the invention; the
 CC against a polypeptide of the invention; a method of detecting
 CC polynucleotides or polypeptides of the invention; an antibody
 CC identifying a compound which binds to a polypeptide; and methods of
 CC ameliorating a medical condition; kits comprising polynucleotide probes
 CC and/or monoclonal antibodies for carrying out the methods of the
 CC invention; methods for the identification of compounds that modulate the
 CC expression or activity of the polynucleotide and/or polypeptide; and 767
 CC contig sequences corresponding to the cDNA sequences of the invention
 CC (ADCC1861-ADCC3262) and the polypeptides encoded by the contigs (ADCC32628

CC -ADCC3394). The nucleic acids and polypeptides of the invention are
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the
 CC identification of mutations responsible for genetic disorders or other
 CC traits, for assessing biodiversity, and in producing many other types of
 CC data and products dependent on DNA and amino acid sequences. They are
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's
 CC disease and other neurodegenerative diseases, anemia, platelet
 CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 CC cancer. The nucleic acids may also be used as hybridization probes or
 CC primers, and in the recombinant production of a protein. The polypeptides
 CC are also useful in generating antibodies, as molecular weight markers,
 CC and as food supplements. The present sequence represents a specifically
 CC claimed human cDNA sequence of the invention. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 2461 BP; 745 A; 438 C; 604 G; 674 T; 0 U; 0 Other;
 CC Query Match 1.6%; Score 21; DB 10; Length 2461;
 CC Best Local Similarity 100.0%; Pred. No. 13;
 CC Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1115 GTGTGTTCCAGTGTCTGTG 1135
 DB 1168 GTGTGTTCCAGTGTCTGTG 1188

RESULT 23
 ADL82850
 ID ADL82850 standard; cDNA; 2562 BP.
 XX AC ADL82850;
 XX XX
 DT 17-JUN-2004 (first entry)
 XX DE Human PRO4912 cDNA, SEQ ID 52.
 XX KW Immunosuppressive; Cytostatic; Antiarthritic; Antirheumatic; Antianemic;
 KW Antiallergic; Muscular; Neuroprotective; Nephroprotective; Antiinflammatory;
 KW Gene therapy; PRO; B cell related disorder; cancer;
 KW Immune-mediated inflammatory disease; human; gene; ss.
 XX OS Homo sapiens.
 XX PN WO2004024097-A2.
 XX PD 25-MAR-2004.
 XX XX
 PF 15-SEP-2003; 2003WO-US029097.
 XX PR 16-SEP-2002; 2002US-0411392P.
 XX PA (GENTH) GENENTECH INC.
 XX PI Chiu H, Clark H, Dennis K, Fong S, Schoenfeld JR, Wood WJ;
 PI Wu TD;
 XX WPI; 2004-329389/30.
 DR P-PSDB; ADL82851.
 XX DE
 XX PT New PRO polypeptide, useful for diagnosing and treating a B cell related
 PT disorder, e.g. Burkitt's lymphoma, rheumatoid arthritis, autoimmune
 PT mediated hemolytic anemia, myasthenia gravis or ankylosing spondylitis.
 XX PT
 XX Claim 2, Fig 52, 695bp; English.
 XX PS
 XX XX
 CC The present invention relates to PRO proteins and their coding sequences.
 CC The PRO proteins are useful for diagnosing and treating a B cell related
 CC disorder, e.g. X-linked infantile hypogammaglobulinemia, polysaccharide
 CC antigen unresponsiveness, selective Iga deficiency, selective IgM
 CC deficiency, selective deficiency of IgG subclasses, immunodeficiency with
 CC hyper IgM, transient hypogammaglobulinemia of infancy, Burkitt's

CC lymphoma, intermediate lymphoma, follicular lymphoma, type II
 CC hypersensitivity, rheumatoid arthritis, autoimmune mediated haemolytic
 CC anaemia, myasthenia gravis, hypoadrenocorticism, glomerulonephritis, or
 CC ankylosing spondylitis. The PRO proteins are also useful for preparing a
 CC medicament for treating a condition that is responsive to the PRO
 CC protein, e.g. cancer or immune-mediated inflammatory diseases. The PRO
 CC coding sequences are useful as hybridization probes in chromosome and
 CC gene mapping, in preparing PRO proteins, or in generating transgenic
 CC animals or knockout animals, which in turn are useful in the development
 CC and screening of therapeutically useful reagents.

CC Sequence 2562 BP; 782 A; 458 C; 623 G; 699 T; 0 U; 0 Other;

Query Match 1.6%; Score 21; DB 12; Length 2562;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGGTGCTGTG 1135
 |||||
 DB 1272 GTGTGTTCCAGGTGCTGTG 1292

RESULT 24

ID ADO19167 standard; cDNA; 2562 BP.

XX ADO19167;

DT 12-AUG-2004 (first entry)

XX Human PRO polynucleotide #51.

XX Human; PRO; gene; ss; immune related disorder;
 KM systemic lupus erythematosus; rheumatoid arthritis; osteoarthritis;
 KM juvenile chronic arthritis; systemic sclerosis; Sjogren's syndrome;
 KM vasculitis; sarcoidosis; autoimmune haemolytic anaemia;
 KM autoimmune thrombocytopenia; thyroditis; diabetes mellitus;
 KM renal disease; demyelinating disease; central nervous system;
 KM peripheral nervous system; demyelinating polynuropathy;
 KM Guillain-Barre syndrome;
 KM chronic inflammatory demyelinating polynuropathy.

XX Homo sapiens.

PN WO2004043361-A2.

PD 27-MAY-2004.

PF 06-NOV-2003; 2003WO-US035268.

PR 08-NOV-2002; 2002US-0425235P.

PA (GETH) GENENTECH INC.

PI Feng S, Dennis K, Clark H, Chiu H, Schoenfeld J, Williams PM;

PI Wood WI, Wu TD;

XX WPI; 2004-420067/39.

DR P-PSDB; ADO19168.

XX Novel PRO polypeptide e.g., PRO69614, PRO71106, or PRO86388 useful for
 PT treating an immune related disorder such as systemic lupus erythematosus,
 PT rheumatoid arthritis, osteoarthritis, juvenile chronic arthritis or
 PT spondyloarthritis.

XX Claim 1; SEQ ID NO 106; 1731bp; English.

XX The invention relates to human PRO polypeptides and the polynucleotides
 CC encoding them. The polypeptides and polynucleotides are useful for
 CC treating and diagnosing immune related disorders in mammals. The immune
 CC related disorders include systemic lupus erythematosus, rheumatoid
 CC arthritis, osteoarthritis, juvenile chronic arthritis, systemic
 CC sclerosis, Sjogren's syndrome, vasculitis, sarcoidosis, autoimmune

CC haemolytic anaemia, autoimmune thrombocytopenia, thyroditis, diabetes
 CC mellitus, immune-mediated renal disease, demyelinating diseases of the
 CC central or peripheral nervous system, demyelinating polynuropathy,
 CC Guillain-Barre syndrome and chronic inflammatory demyelinating
 CC polynuropathy. This sequence represents a human PRO polynucleotide of
 CC the invention.

CC Sequence 2562 BP; 782 A; 458 C; 623 G; 699 T; 0 U; 0 Other;

Query Match 1.6%; Score 21; DB 12; Length 2562;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGGTGCTGTG 1135
 |||||
 DB 1272 GTGTGTTCCAGGTGCTGTG 1292

RESULT 25

ID ADO60113 standard; DNA; 2562 BP.

XX ADO60113;

DT 26-AUG-2004 (first entry)

XX Human CCR6 gene SEQ ID NO:1.

XX de; human; gene; chaperonin containing T-complex 1 subunit 6; CCR6;

KM retinoblastoma pathway; cytoskeletal.

XX Homo sapiens.

FT Key Location/Qualifiers

FT CDS 56..1651

FT /*tag= a

FT /product= "CCR6"

PN WO2004048541-A2.

PD 10-JUN-2004.

PF 24-NOV-2003; 2003WO-US037548.

PR 25-NOV-2002; 2002US-0428872P.

PA (EXEL-) EXELIXIS INC.

PI Tai AK, Song C, Ollmann MM, Gillett LA, Lickteig K;

PI WPI; 2004-441173/41.

DR P-PSDB; ADO60120.

XX Use of Chaperonin containing T-complex 1 subunit 6 (CCR6) polypeptides or

PT nucleic acids, e.g. identifying a candidate retinoblastoma (RB) pathway

PT modulating agent or modulating a RB pathway of a cell.

XX Example 1; SEQ ID NO 1; 57bp; English.

XX The invention relates to a novel use of Chaperonin containing T-complex 1
 CC subunit 6 (CCR6) polypeptides or nucleic acids for identifying a
 CC candidate retinoblastoma (RB) pathway modulating agent or for modulating
 CC a RB pathway of a cell. An RB pathway modulating agent of the invention
 CC has cytostatic activity. The CCR6 polypeptides or nucleic acids are
 CC useful for identifying a candidate RB pathway modulating agent or for
 CC modulating a RB pathway of a cell. The CCR6-modulating agents are useful
 CC as therapeutic agents for treating disorders associated with defective or
 CC impaired RB function and/or CCR6 function. CCR6 interacting proteins are
 CC also useful in detecting and providing information about the function of
 CC CCR6 proteins. The present sequence represents a CCR6 nucleic acid of the
 CC invention.

CC Sequence 2562 BP; 782 A; 458 C; 623 G; 699 T; 0 U; 0 Other;

Query Match 1.6%; Score 21; DB 12; Length 2562;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGGTTCAGGTCTGGTG 1135
 DB 1272 GTGTGGTTCAGGTCTGGTG 1292

RESULT 26
 ADO60114
 ID ADO60114 standard; DNA; 2562 BP.

AC ADO60114;
 DT 26-AUG-2004 (first entry)
 DE Human CCR6 gene SEQ ID NO:2.
 XX
 KW de; human; gene; chaperonin containing T-complex 1 subunit 6; CCT6;
 KW retinoblastoma pathway; cytosstatic.
 OS Homo sapiens.

Key Location/Qualifiers
 CDS 56..1651
 FT /*tag= a
 FT /product= "CCT6"

PN WO2004048541-A2.
 PD 10-JUN-2004.

PF 24-NOV-2003; 2003WO-US037548.

PR 25-NOV-2002; 2002US-0428872P.

PA (EXEL-) EXELIXIS INC.

PI Tai AK, Song C, Ollmann MM, Gillelt LA, Lickteig K;

DR WPI: 2004-441173/41.
 DR P-PSDB: ADO60120.

XX Use of Chaperonin containing T-complex 1 subunit 6 (CCT6) polypeptides or
 PT nucleic acids, e.g. identifying a candidate retinoblastoma (RB) pathway
 PT modulating agent or modulating a RB pathway of a cell.

PS Disclosure; SEQ ID NO 2; 57pp; English.

XX The invention relates to a novel use of Chaperonin containing T-complex 1
 CC subunit 6 (CCT6) polypeptides or nucleic acids for identifying a
 CC candidate retinoblastoma (RB) pathway modulating agent or for modulating
 CC a RB pathway of a cell. An RB pathway modulating agent of the invention
 CC has cytosstatic activity. The CCT6 polypeptides or nucleic acids are
 CC useful for identifying a candidate RB pathway modulating agent or for
 CC modulating a RB pathway of a cell. The CCT6-modulating agents are useful
 CC as therapeutic agents for treating disorders associated with defective or
 CC impaired RB function and/or CCT6 function. CCT6 interacting proteins are
 CC also useful in detecting and providing information about the function of
 CC CCT6 proteins. The present sequence represents a CCT6 nucleic acid of the
 CC invention.

XX Sequence 2562 BP; 782 A; 458 C; 623 G; 699 T; 0 U; 0 Other;

Query Match 1.6%; Score 21; DB 12; Length 2562;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGGTTCAGGTCTGGTG 1135
 DB 1272 GTGTGGTTCAGGTCTGGTG 1292

RESULT 27

AAFI5841
 ID AAFI5841 standard; cDNA; 2594 BP.

AC AAFI5841;
 DT 13-MAR-2001 (first entry)

DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:276.

XX
 KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
 KW neuroprotective; cytosstatic; cardioactive; immunomodulatory; muscular;
 KW vulnary; gastrointestinal; nephrotoxic; antiinfective; gynaecological;
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
 XX wound; infectious disease; ss.

OS Homo sapiens.

PN WO200055174-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US005988.

PR 12-MAR-1999; 99US-0124270P.

PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.

PI Rosen CA, Ruben SM;

DR WPI: 2000-587513/55.

DR P-PSDB: AAB56638.

XX Prostate cancer associated gene sequences, referred to as prostate cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as prostate cancer.

PS Claim 1; Page 796; 2338pp; English.

XX AAFI566 to AAFI6505 encode the human prostate cancer associated
 CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
 CC The prostate cancer antigens can have neuroprotective, cytosstatic,
 CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
 CC nephrotoxic, antiinfective, gynaecological and antibacterial activities,
 CC and can be used in gene therapy. The prostate cancer antigen
 CC polynucleotides may be used for detection of prostate cancer, chromosome
 CC identification, as chromosome markers, and for numerous other diagnostic
 CC or research purposes. The prostate cancer antigens may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders, wounds, and infectious diseases. AAFI566 to AAFI6514 to
 CC AAB57303 represent sequences used in the exemplification of the present
 CC invention

XX Sequence 2594 BP; 789 A; 473 C; 628 G; 703 T; 0 U; 1 Other;

Query Match 1.6%; Score 21; DB 3; Length 2594;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGGTTCAGGTCTGGTG 1135
 DB 1298 GTGTGGTTCAGGTCTGGTG 1318

RESULT 28

ADO60115
 ID ADO60115 standard; DNA; 2647 BP.

XX

```

AC ADO60115;
XX
XX 26-AUG-2004 (first entry)
XX
XX Human CCT6 gene SEQ ID NO:3.
XX
XX de; human; gene; chaperonin containing T-complex 1 subunit 6; CCT6;
XX retinoblastoma pathway; cytosolic.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 137..1732
XX /*tag= a
XX /*product= "CCT6"
XX /*transl_except= (pos:1037..1039,aa:Ser)
XX /*transl_except= (pos:1043..1045,aa:Asp)
XX
XX MO2004048541-A2.
XX
XX 10-JUN-2004.
XX
XX 24-NOV-2003; 2003MO-US037548.
XX
XX 25-NOV-2002; 2002US-0428872P.
XX
XX (EXEL-) EXELIXIS INC.
XX
XX Tai AK, Song C, Ollmann MM, Gillett LA, Lickteig K;
XX WPI; 2004-441173/41.
XX P-PSDB; ADO60120.
XX
XX Use of Chaperonin containing T-complex 1 subunit 6 (CCT6) polypeptides or
XX nucleic acids, e.g. identifying a candidate retinoblastoma (RB) pathway
XX modulating agent or modulating a RB pathway of a cell.
XX
XX Disclosure; SEQ ID NO 3; 57pp; English.
XX
XX The invention relates to a novel use of Chaperonin containing T-complex 1
XX subunit 6 (CCT6) polypeptides or nucleic acids for identifying a
XX candidate retinoblastoma (RB) pathway modulating agent or for modulating
XX a RB pathway of a cell. An RB pathway modulating agent of the invention
XX has cytosolic activity. The CCT6 polypeptides or nucleic acids are
XX useful for identifying a candidate RB pathway modulating agent or for
XX modulating a RB pathway of a cell. The CCT6-modulating agents are useful
XX as therapeutic agents for treating disorders associated with defective or
XX impaired RB function and/or CCT6 function. CCT6 interacting proteins are
XX also useful in detecting and providing information about the function of
XX CCT6 proteins. The present sequence represents a CCT6 nucleic acid of the
XX invention.
XX
XX Sequence 2647 BP; 796 A; 487 C; 654 G; 710 T; 0 U; 0 Other;
SQ

```

```

Query Match 1.6%; Score 21; DB 12; Length 2647;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1115 GTGTGTTCCAGTGCTGTG 1135
DB 1353 GTGTGTTCCAGTGCTGTG 1373

```

```

RESULT 29
ABV24814
ID ABV24814 standard; cDNA; 2837 BP.
XX
XX ABV24814;
AC
XX
XX 16-SEP-2002 (first entry)
DT
XX
XX Human prostate expression marker cDNA 24805.
DE
XX
XX

```

```

KM Human; prostate cancer; cytosolic; carcinogen; pharmacodynamic marker;
KM pharmacogenomic marker; gene; se.
XX
XX Homo sapiens.
XX
XX WO200160860-A2.
XX
XX 23-AUG-2001.
XX
XX 20-FEB-2001; 2001MO-US005171.
XX
XX 17-FEB-2000; 2000US-0183319P.
XX 16-MAR-2000; 2000US-0189662P.
XX 25-MAY-2000; 2000US-0207454P.
XX 09-JUN-2000; 2000US-0211314P.
XX 18-JUL-2000; 2000US-0219007P.
XX 13-DEC-2000; 2000US-0255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Endege WO, Morahan JE;
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX Claim 1; Page 4756; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX specification or its complement. (I) is useful for: (a) assessing whether
XX a patient is afflicted with prostate cancer; (b) monitoring the
XX progression of prostate cancer in a patient; (c) assessing the efficacy
XX of a test compound to inhibit prostate cancer in a patient; (d) assessing
XX the efficacy of a therapy for inhibiting prostate cancer in a patient;
XX (e) selecting a composition for inhibiting prostate cancer in a patient;
XX (f) assessing the prostate cell carcinogenic potential of a compound; (g)
XX determining whether prostate cancer has metastasized in a patient; (h)
XX assessing the aggressiveness or indolence of prostate cancer in a patient
XX ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
XX Sequence 2837 BP; 855 A; 538 C; 680 G; 764 T; 0 U; 0 Other;
SQ

```

```

Query Match 1.6%; Score 21; DB 5; Length 2837;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1115 GTGTGTTCCAGTGCTGTG 1135
DB 1307 GTGTGTTCCAGTGCTGTG 1327

```

```

RESULT 30
AAK79798
ID AAK79798 standard; DNA; 16682 BP.
XX
XX AAK79798;
AC
XX
XX 07-NOV-2001 (first entry)
DT
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34610.
DE
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytosolic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX

```

PR 17-JAN-2001; 2001WC-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217486P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0228287P.
PR 01-SEP-2000; 2000US-0228343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236377P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.

PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246533P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251989P.
PR 06-DEC-2000; 2000US-0251479P.
PR 06-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
DR
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,

PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX Disclosure; SEQ ID NO 34610; 3071bp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 16682 BP; 4217 A; 3754 C; 4167 G; 4538 T; 0 U; 6 Other;

Query Match 1.6%; Score 21; DB 4; Length 16682;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 CTTGCTGTGGATTTGGTT 57
DB 1920 CTTGCTGTGGATTTGGTT 1940

RESULT 31
AAK70010
ID AAK70010 standard; DNA; 16682 BP.
XX
AC AAK70010;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24822.
XX
KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; de.
OS Homo sapiens.
XX
PN WO200157182-A2.
PD
XX 09-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189674P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231245P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234988P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235534P.
PR 27-SEP-2000; 2000US-0235636P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239355P.
PR 13-OCT-2000; 2000US-0239393P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241212P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-02446474P.
PR 08-NOV-2000; 2000US-02446475P.
PR 08-NOV-2000; 2000US-02446476P.
PR 08-NOV-2000; 2000US-02446477P.

PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241866P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.

PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251866P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251889P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Disclosure; SEQ ID NO 34611; 3071pp + Sequence Listing; English.
XX
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytosstatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX CC proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX CC supplement the patients own production of (I). Additionally, (I)
XX CC polynucleotides may be used to produce the secreted (I), by inserting the
XX nucleic acids into a host cell and culturing the cell to express the
XX CC protein. (I) proteins and polynucleotides may be used to prevent,
XX CC diagnose and treat immune/haematopoietic-related diseases, especially
XX CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX CC to AAK87694 represent human immune/haematopoietic antigen genomic
XX CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX CC represent sequences used in the exemplification of the present invention
XX
XX Sequence 21436 BP; 5267 A; 4950 C; 5418 G; 5801 T; 0 U; 0 Other;
XX
XX
XX Query Match 1.6%; Score 21; DB 4; Length 21436;
XX Best Local Similarity 100.0%; Pred. No. 12;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 37 CTTGCTGTGGATATTGGTT 57
XX |||||||
XX Db 1920 CTTGCTGTGGATATTGGTT 1940
XX
XX
XX RESULT 33
XX AAK70011
XX ID AAK70011 standard; DNA; 21436 BP.
XX
XX AAK70011;
XX
XX AC AAK70011;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24823.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW Cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
PN WO200157182-A2.
PD 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US001354.
PF
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220966P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 23-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0232081P.
PR 14-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.

PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241825P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249255P.
PR 17-NOV-2000; 2000US-0249257P.
PR 17-NOV-2000; 2000US-0249259P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251866P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-0251990P.

PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 P1 Rosen CA, Barash SC, Ruben SM;
 DR WPI; 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.
 XX
 PS Disclosure; SEQ ID NO 24823; 3071bp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I) additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK81694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention, AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 21436 BP; 5267 A; 4950 C; 5418 G; 5801 T; 0 U; 0 Other;

	Query Match Similarity	1.6%	Score 21;	DB 4;	Length 21336;
	Best Local Similarity	100.0%;	Pred. No. 12;		
	Matches	21;	Conservative	0;	Mismatches
				0;	Indels
				0;	Gaps
Oy	37	CTTGCTGATGGATTATTTGGTT	57		
Db	1920	CTTGCCTGGATGGATTATTTGGTT	1940		

RESULT 34
AAD02795
ID AAD02795 standard; DNA; 33 BP.

AC AAD02795;

DT 31-MAY-2001 (first entry)

DE Candida albicans ERG8 coding sequence amplifying primer #1.

KM Phosphomevalonate kinase; PMK; ERG8; anti-fungal agent; diagnosis
KM infection; PCR primer; BS.

OS **Candida albicans.**

PN WO200114533-A2.

01-MAR-2001. PD

PF 15-AUG-2000; 2000WO-GB003100.

PR 21-AUG-1999; 99GB-00019766.

PA (ASTR) ASTRAZENECA AB.
PA (ASTR) ASTRAZENECA UK LTD.

PI Rosamond JDC, Schnell NF;

DR WPI; 2001-218441/22.

X

PT New polypeptides and polynucleotides (ERG8) from *Candida albicans*, useful
PT in assays for identifying inhibitors of phosphomethylaldehyde kinase activity
PT and as reagents for diagnosing *C. albicans* infection.

PS Example 4; Page 29; 29pp; English.

The patent discloses phosphomevalonate kinase (PMK; ERG8) protein and the corresponding DNAs from candida albicans. The ERG8 protein is useful in assays for identifying compounds that inhibit phosphomevalonate kinase (PMK) activity. These inhibitors are useful as antifungal agents. The ERG8 DNA and protein are also useful as reagents for diagnosing C. albicans infection. The present DNA sequence is PCR primer which is used to amplify the Candida albicans ERG8 coding sequence. This sequence incorporates restriction enzyme sites in the ERG8 coding sequence and facilitate its cloning

Sequence 33 BP; 10 A; 8 C; 7 G; 8 T; 0 U; 0 Other;

Query Match	Score	DB	Length
1.5%	20	4	33

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGTCAAAGCATTAGTGC 20

Db 14 ATGTCAAAAGCATTAGTGC 33

1
1
2
1
3
2
1

ACH70211/c

XX
20150611

XX 30 TH 3004 (E: 2nd 3rd 4th)

XX
XX

XX
XX

KW alternative splicing event; genomic alteration

OS Homo sapiens.

PN US2003194704-A1

PD 16-OCT-2003.

03-APR-2002; 2002US-00029386.

PR 03-APR-2002; 2002US-00029386.

PA (PENN/) PENN S G.
PA (PAWR/) PAWR D P

PA (HANZ/) HANZEL D K.

PI Penn SG, Rank DR, Hanzel DK, vv

DR WPI; 2004-119264/12
yy

PT New human genome-derived single exon nucleic acid probes useful for human gene identification and characterization

PT splicing events, for assessing genomic alterations or as tools for

XX
XX
CJ 34w 1E. GEO TD N

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

The invention relates to a nucleic acid probe for measuring human gene expression, comprising any of the 27,400 fully defined nucleotide sequences in the specification, or their complements or fragments, and encoding at least 8 amino acids of any of the 6888 amino acid sequences fully defined in the specification. The probe is a single exon probe that hybridises under high stringency conditions to a nucleic acid molecule expressed in human cells or tissues. Also included are a spatially-addressable set of single exon nucleic acid probes for measuring human

CC gene expression (comprising a plurality of single exon nucleic acid
CC probes cited above, where each of the plurality of probes is separately
CC and addressably isolatable or amplifiable from the plurality), a single
CC exon microarray for measuring human gene expression, a method of
CC measuring human gene expression, a vector comprising the single exon
CC probe cited above, an ORF-encoded peptide comprising at least 8
CC contiguous amino acids of any of the above-mentioned amino acid
CC sequences (optionally with conservative amino acid substitutions), an
CC isolated antibody that binds specifically to a peptide cited above,
CC a method of selling and/or licensing single exon probes or microarrays to
CC a customer desiring to measure gene expression, a method of providing
CC storage medium which contains a database having a plurality of records
CC (each record including data on the expression of a single exon probe
CC cited above). The probe, method and apparatus are useful in gene
CC expression analysis. The probes may be used as tools for surveying
CC tissues to detect the presence of expressed messages that contain their
CC specific exon, or in constructing genome-derived single exon microarrays.
CC In addition, the probes are used in identifying and characterizing
CC alternative splicing events, in detecting and characterizing gross
CC alterations in the genomic locus that includes their exon, in assessing
CC smaller genomic alterations, in priming the synthesis of nucleic acids,
CC or in expressing the ORF-encoded peptide. Note: The present sequence is a human
CC single exon probe of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docID=20030194704

SQ Sequence 502 BP; 122 A; 148 C; 121 G; 111 T; 0 U; 0 Other;

Query Match 1.5%; Score 20; DB 12; Length 502;
Best Local Similarity 100.0%; Pred. No. 43;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 723 CTGGGAATTCACATGAAA 742
|||
Db 95 CTGGGAATTCACATGAAA 76

RESULT 36

AA181760
ID AA181760 standard; cDNA; 541 BP.

AC AA181760;

DT 06-NOV-2001 (first entry)

DB Human polynucleotide SEQ ID NO 1820.

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.

OS Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US004927.

PR 28-FEB-2000; 2000US-00515126.

PR 18-MAY-2000; 2000US-00577409.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI; 2001-514838/56.

DR P-PSDB; AAC01829.

PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing

PT and treating e.g. leukemia, inflammation and immune disorders.
XX
PS Claim 1; SEQ ID NO 1820; 1399bp + Sequence Listing; English.

CC The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibit activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 541 BP; 151 A; 63 C; 71 G; 110 T; 0 U; 146 Other;

Query Match 1.5%; Score 20; DB 4; Length 541;
Best Local Similarity 100.0%; Pred. No. 43;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1019 TACAAGCATTACACAAAA 1038
|||
Db 479 TACAAGCATTACACAAAA 498

RESULT 37

AB212796
ID AB212796 standard; DNA; 630 BP.

AC AB212796;

DT 21-JAN-2003 (first entry)

DE Arabidopsis thaliana stress regulated gene SEQ ID NO 601.

KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.

OS Arabidopsis thaliana.

PN WO200216655-A2.

PD 28-FEB-2002.

PF 24-AUG-2001; 2001WO-US026685.

PR 24-AUG-2000; 2000US-0227866P.

PR 26-JAN-2001; 2001US-0264647P.

PR 22-JUN-2001; 2001US-0300111P.

PA (SCRI) SCRIPPS RES INST.

PI Harper JF, Kreps J, Wang X, Zhu T;

DR WPI; 2002-304127/34.

PT Identifying a stress condition to which a plant cell has been exposed and
PT producing plants with increased tolerance to these abiotic stresses.

PS Claim 144; SEQ ID NO 601; 577bp + Sequence Listing; English.

CC The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising: (a) contacting nucleic acid
CC representative of expressed polynucleotides in the plant cell with an
CC array or probes representative of the plant cell genome; and (b)
CC detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that

CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used
CC in methods of the invention. Note: The sequence data for this patent is
CC not represented in the printed specification but is based on sequence
CC information supplied to Derwent by the European Patent Office
XX
SQ Sequence 630 BP, 162 A, 136 C, 142 G, 190 T, 0 U, 0 Other;

Query Match 1.5%; Score 20; DB 6; Length 630;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 CTGCTGATATTGTTCTT 60
|||||
DB 317 CTGCTGATATTGTTCTT 336

RESULT 38

AB076676/c
ID AB076676 standard; DNA; 46852 BP.

AC AB076676;

XX 26-MAR-2003 (first entry)

XX Androgen receptor signalling pathway-associated DNA AH007803.

XX Androgen receptor; transactivation; modulator; Smad3; Smad4; Akt; TGF-B;

XX signal transduction pathway; transforming growth factor-B; phosphatase;

XX tensin; cytoskeletal; antiproliferative; cellular proliferation; cancer;

XX AH007803; ds.

XX Synthetic.

XX WO200282081-A2.

XX 17-OCT-2002.

XX 05-APR-2002; 2002WO-US011086.

XX 06-APR-2001; 2001US-0282266P.

XX 13-MAR-2002; 2002US-0365060P.

XX (UVRP) UNIV ROCHESTER.

XX Chang C;

XX WPI; 2003-046871/04.

XX Modulating androgen receptor activity, by administering a compound that

XX modulates receptor activity, inhibits receptor-signal transduction

XX pathway/receptor-coactivator interaction or changes amount or receptor.

XX Disclosure; Page 227-240; 302pp; English.

XX This invention describes a novel method for modulating androgen receptor

XX activity or androgen receptor-mediated transactivation activity in a

XX cell. The method involves administering a compound which causes

XX modulation of the androgen receptors activity and the inhibition of

XX interaction between the receptor and a protein involved in a signal

XX transduction pathway. The compound also inhibits the interaction between

XX the androgen receptor and a protein selected from Smad3, Smad4, Akt,

XX transforming growth factor (TGF)-B and phosphatase and tensin homologues

XX deleted on chromosome 10 (PTEN) or their fragments. The compounds of the

XX invention have cytostatic and antiproliferative activity. The obtained

XX composition is useful for treating any disease, where uncontrolled

XX proliferation or cellular proliferation occurs such as cancer, e.g.

XX prostate cancer. This sequence represents the androgen receptor

XX transactivation signalling pathway modulator AH007803 described in the

XX method of the invention

XX Sequence 46852 BP; 13662 A; 7976 C; 9059 G; 15917 T; 0 U; 238 Other;

XX Query Match 1.5%; Score 20; DB 8; Length 46852;

Best Local Similarity 100.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 139 AAAGATCTAGATCAAAAT 158
|||||
DB 17354 AAAGATCTAGATCAAAAT 17335

RESULT 39
AB076678/c
ID AB076678 standard; DNA; 218336 BP.

AC AB076678;

XX 26-MAR-2003 (first entry)

XX Androgen receptor signalling pathway-associated DNA AF067844.

XX Androgen receptor; transactivation; modulator; Smad3; Smad4; Akt; TGF-B;

XX signal transduction pathway; transforming growth factor-B; phosphatase;

XX tensin; cytoskeletal; antiproliferative; cellular proliferation; cancer;

XX AF067844; ds.

XX Synthetic.

XX WO200282081-A2.

XX 17-OCT-2002.

XX 05-APR-2002; 2002WO-US011086.

XX 06-APR-2001; 2001US-0282266P.

XX 13-MAR-2002; 2002US-0365060P.

XX (UVRP) UNIV ROCHESTER.

XX Chang C;

XX WPI; 2003-046871/04.

XX This invention describes a novel method for modulating androgen receptor

XX activity or androgen receptor-mediated transactivation activity in a

XX cell. The method involves administering a compound which causes

XX modulation of the androgen receptors activity and the inhibition of

XX interaction between the receptor and a protein involved in a signal

XX transduction pathway. The compound also inhibits the interaction between

XX the androgen receptor and a protein selected from Smad3, Smad4, Akt,

XX transforming growth factor (TGF)-B and phosphatase and tensin homologues

XX deleted on chromosome 10 (PTEN) or their fragments. The compounds of the

XX invention have cytostatic and antiproliferative activity. The obtained

XX composition is useful for treating any disease, where uncontrolled

XX proliferation or cellular proliferation occurs such as cancer, e.g.

XX prostate cancer. This sequence represents the androgen receptor

XX transactivation signalling pathway modulator AF067844 described in the

XX method of the invention

XX Sequence 218336 BP; 64194 A; 39437 C; 43295 G; 71406 T; 0 U; 4 Other;

XX Query Match 1.5%; Score 20; DB 8; Length 218336;

XX Best Local Similarity 100.0%; Pred. No. 34;

XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 139 AAAGATCTAGATCAAAAT 158
|||||
DB 82361 AAAGATCTAGATCAAAAT 82342

RESULT 40

ABL88534
ID ABL88534 standard; DNA; 355 BP.

XX
AC ABL88534;

XX
DT 24-MAY-2002 (first entry)

XX
DE C marmoreus mu-conopeptide Mr3-3 coding sequence.

XX
KM Mu-conopeptide; snail; venom; voltage-gated sodium channel; analgesic;

XX
KM anaesthetic; neuromuscular blocking agent; neuroprotective; pain;

XX
KM cerebroprotective; anticonvulsant; antiaging; antidiabetic;

XX
KM cardiovascular; vasotropic; cardiac; tranquilizer; antimigraine;

XX
KM neurodegenerative disease; neuromuscular disorder; gene; ds.

XX
OS Cornus marmoreus.

XX
PN WO200207678-A2.

XX
PD 31-JAN-2002.

XX
PF 23-JUL-2001; 2001WO-US023125.

XX
PR 21-JUL-2000; 2000US-0219619P.

XX
PR 03-NOV-2000; 2000US-0245157P.

XX
PR 29-JAN-2001; 2001US-0264319P.

XX
PR 21-MAR-2001; 2001US-0277270P.

XX
PA (UTAH) UNIV UTAH RES FOUND.

XX
PA (COGN-) COGNETIX INC.

XX
PI Oliveira BM, McIntosh JM, Garrett JE, Watkins M, Cruz LJ, Shon K;

XX
PI Jacobsen R, Jones RM, Cartier GE, Shen GS,

XX
DR WPI, 2002-217020/27.

XX
DR P-PSDB; ABB88246.

XX
PT New mu-conopeptides useful for treating disorders associated with voltage

XX
PT agents, as local anesthetic agents, as analgesic agents and as

XX
PT neuroprotective agents.

XX
PS Claim 8; Page 48; 231pp; English.

XX
CC The present invention relates to mu-conopeptides derived from snails,

XX
CC which can be in the treatment of disorders associated with voltage-gated

XX
CC ion channels. These may include neurodegenerative disorders such as

XX
CC amyotrophic lateral sclerosis, neuromuscular disorders, post-operative or

XX
CC severe chronic pain, stroke, trigeminal neuralgia, diabetic neuropathy,

XX
CC post-herpetic neuralgia, neuronal pain and phantom limb, burn pain,

XX
CC epilepsy, for reducing neurotoxic injury associated with hypoxia, anoxia

XX
CC or ischaemia which typically follows stroke, cerebrovascular accident,

XX
CC brain or spinal cord trauma, myocardial infarct, physical trauma,

XX
CC drowning, suffocation, perinatal asphyxia, or hypoglycaemic events, for

XX
CC providing muscle relaxation, treating essential blepharospasm and other

XX
CC forms of focal dystonia, and for anti-wrinkle use. The present sequence

XX
SQ is a mu-conopeptide preprotein coding sequence of the invention

XX
SQ Sequence 355 BP; 72 A; 93 C; 91 G; 99 T; 0 U; 0 Other;

XX
SQ Query Match 1.5%; Score 19; DB 6; Length 355;

XX
SQ Best Local Similarity 100.0%; Pred. No. 1.3e+02;

XX
SQ Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX
QY 861 TGTGTGATGACGCTT 879

XX
DB 262 TGTGTGATGACGCTT 280

XX
Search completed: January 26, 2005, 08:04:55

XX
Job time : 536 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 26, 2005, 09:09:55 ; Search time 552 Seconds
(without alignments)
13521.574 Million cell updates/sec

Title: US-10-069-062-6

Perfect score: 1299

Sequence: 1 atgcctaaagcatttagtgc.....aagactatcatagttcataa 1299

Scoring table: OLIGO NUC
Gapex 60.0, Gapext 60.0

Searched: 4300275 seqs, 2872944193 residues

Word size: 0

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database:

1: Published Applications NA.*
2: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_NEW_PUB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	891	68.6	1299	US-10-033-585-6023	Sequence 6023, Ap
2	65	5.0	65	US-10-033-585-3023	Sequence 3023, Ap
3	22	1.7	22	US-10-033-585-5023	Sequence 5023, Ap
4	22	1.7	690	US-10-027-632-150718	Sequence 150718, Ap
5	22	1.7	690	US-10-027-632-150718	Sequence 150718, Ap
6	22	1.7	711	US-10-027-632-26202	Sequence 26202, A
7	22	1.7	711	US-10-027-632-26202	Sequence 26202, A
8	22	1.7	711	US-10-027-632-26202	Sequence 26202, A
9	22	1.7	711	US-10-027-632-26202	Sequence 26202, A
10	21	1.6	313	US-10-242-535A-6452	Sequence 6452, Ap
11	21	1.6	313	US-10-085-783A-6452	Sequence 6452, Ap
12	21	1.6	400	US-10-767-795-5310	Sequence 5310, Ap

13	21	1.6	476	US-09-918-995-17749	Sequence 17749, A
14	21	1.6	479	US-09-920-300A-726	Sequence 726, Ap
15	21	1.6	479	US-10-033-528-726	Sequence 726, Ap
16	21	1.6	479	US-10-099-926-726	Sequence 726, Ap
17	21	1.6	530	US-09-998-598-263	Sequence 263, Ap
18	21	1.6	530	US-09-920-300A-1254	Sequence 1254, Ap
19	21	1.6	530	US-10-033-528-1254	Sequence 1254, Ap
20	21	1.6	530	US-10-099-926-1254	Sequence 1254, Ap
21	21	1.6	1060	US-10-198-846-10192	Sequence 10192, A
22	21	1.6	1574	US-10-264-049-594	Sequence 594, Ap
23	21	1.6	1825	US-10-723-860-5689	Sequence 5689, Ap
24	21	1.6	2010	US-10-170-385-414	Sequence 414, Ap
25	21	1.6	2010	US-10-172-118-662	Sequence 662, Ap
26	21	1.6	2010	US-10-159-553-933	Sequence 393, Ap
27	21	1.6	2010	US-10-342-887-662	Sequence 662, Ap
28	21	1.6	2594	US-09-925-300-276	Sequence 276, Ap
29	21	1.6	2837	US-10-357-930-24803	Sequence 24803, A
30	21	1.5	371	US-09-796-692-9039	Sequence 9039, Ap
31	20	1.5	371	US-10-040-862-9039	Sequence 9039, Ap
32	20	1.5	371	US-10-057-475B-9039	Sequence 9039, Ap
33	20	1.5	371	US-10-154-884B-9039	Sequence 9039, Ap
34	20	1.5	371	US-10-764-374-9039	Sequence 3406, Ap
35	20	1.5	502	US-10-029-386-3406	Sequence 3406, Ap
36	20	1.5	630	US-09-938-842A-601	Sequence 601, Ap
37	20	1.5	630	US-09-938-842A-601	Sequence 601, Ap
38	20	1.5	728	US-10-027-632-160412	Sequence 160412, A
39	20	1.5	728	US-10-027-632-160412	Sequence 160412, A
40	20	1.5	1656	US-10-369-493-43499	Sequence 43499, A
41	20	1.5	46852	US-10-473-939-8	Sequence 8, Appl
42	20	1.5	218336	US-10-473-939-10	Sequence 10, Appl
43	19	1.5	233	US-10-424-559-47771	Sequence 47771, A
44	19	1.5	287	US-10-425-115-143271	Sequence 143271, A
45	19	1.5	355	US-09-910-009A-157	Sequence 157, Ap
46	19	1.5	355	US-10-828-478-157	Sequence 157, Ap
47	19	1.5	538	US-10-027-632-91442	Sequence 91442, A
48	19	1.5	538	US-10-027-632-91443	Sequence 91443, A
49	19	1.5	538	US-10-027-632-91444	Sequence 91444, A
50	19	1.5	538	US-10-027-632-91445	Sequence 91445, A
51	19	1.5	538	US-10-027-632-91446	Sequence 91446, A
52	19	1.5	538	US-10-027-632-91447	Sequence 91447, A
53	19	1.5	538	US-10-027-632-91448	Sequence 91448, A
54	19	1.5	538	US-10-027-632-91449	Sequence 91449, A
55	19	1.5	542	US-10-027-632-291867	Sequence 291867, A
56	19	1.5	542	US-10-027-632-291868	Sequence 291868, A
57	19	1.5	542	US-10-027-632-291869	Sequence 291869, A
58	19	1.5	542	US-10-027-632-291870	Sequence 291870, A
59	19	1.5	593	US-10-313-542-59	Sequence 89, Appl
60	19	1.5	628	US-10-027-632-270097	Sequence 270097, A
61	19	1.5	628	US-10-027-632-270097	Sequence 270097, A
62	19	1.5	639	US-10-027-632-261652	Sequence 261652, A
63	19	1.5	639	US-10-027-632-261652	Sequence 261652, A
64	19	1.5	662	US-10-027-632-237370	Sequence 237370, A
65	19	1.5	662	US-10-027-632-237371	Sequence 237371, A
66	19	1.5	662	US-10-027-632-237370	Sequence 237370, A
67	19	1.5	662	US-10-027-632-237371	Sequence 237371, A
68	19	1.5	714	US-10-767-701-19874	Sequence 19874, A
69	19	1.5	842	US-10-027-632-162890	Sequence 162890, A
70	19	1.5	842	US-10-027-632-162891	Sequence 162891, A
71	19	1.5	842	US-10-027-632-162892	Sequence 162892, A
72	19	1.5	842	US-10-027-632-162893	Sequence 162893, A
73	19	1.5	842	US-10-027-632-162891	Sequence 162891, A
74	19	1.5	842	US-10-027-632-162892	Sequence 162892, A
75	19	1.5	858	US-10-027-632-10012	Sequence 10012, A
76	19	1.5	858	US-10-027-632-10012	Sequence 10012, A
77	19	1.5	870	US-10-369-493-11994	Sequence 41994, A
78	19	1.5	894	US-10-437-963-37487	Sequence 37487, A
79	19	1.5	995	US-10-363-345A-2114	Sequence 2114, Ap
80	19	1.5	995	US-10-363-345A-2114	Sequence 2114, Ap
81	19	1.5	1111	US-10-425-114-10780	Sequence 10780, A
82	19	1.5	1140	US-10-425-115-182260	Sequence 280260, A
83	19	1.5	1157	US-10-425-114-28009	Sequence 28009, A
84	19	1.5	1231	US-10-437-963-37483	Sequence 37483, A
85	19	1.5	1275	US-10-425-115-87852	Sequence 87852, A


```
662 16 1.2 560 15 US-10-029-386-3425 Sequence 3425, Ap
663 16 1.2 560 15 US-10-027-632-93845 Sequence 93845, A
964 16 1.2 560 15 US-10-027-632-93845 Sequence 93846, A
965 16 1.2 560 15 US-10-027-632-305252 Sequence 305252,
966 16 1.2 560 15 US-10-027-632-305253 Sequence 305253,
967 16 1.2 561 13 US-10-027-632-290308 Sequence 290308,
968 16 1.2 561 13 US-10-027-632-290308 Sequence 290308,
969 16 1.2 561 18 US-10-027-632-290308 Sequence 290308,
970 16 1.2 562 17 US-10-027-632-240337 Sequence 240337,
971 16 1.2 564 13 US-10-027-632-240337 Sequence 4032, Ap
972 16 1.2 564 15 US-10-029-386-4032 Sequence 240337,
973 16 1.2 564 15 US-10-027-632-240337 Sequence 4031, Ap
974 16 1.2 564 17 US-10-027-632-240337 Sequence 20626, A
975 16 1.2 565 17 US-10-029-386-2686 Sequence 7731, Ap
976 16 1.2 571 9 US-09-864-761-7731 Sequence 2686, Ap
977 16 1.2 571 15 US-10-029-386-2686 Sequence 18, Appl
978 16 1.2 572 15 US-10-029-386-2686 Sequence 5388, Ap
979 16 1.2 572 15 US-10-029-386-2686 Sequence 93, Appl
980 16 1.2 573 16 US-10-029-386-2686 Sequence 66415, A
981 16 1.2 576 9 US-09-764-868-523 Sequence 523, App
982 16 1.2 576 15 US-10-029-386-4643 Sequence 4643, Ap
983 16 1.2 577 13 US-10-027-632-322970 Sequence 322970,
984 16 1.2 577 13 US-10-027-632-322971 Sequence 322971,
985 16 1.2 577 15 US-10-027-632-322971 Sequence 322970,
986 16 1.2 577 15 US-10-027-632-322971 Sequence 322971,
987 16 1.2 579 13 US-10-027-632-289790 Sequence 289790,
988 16 1.2 579 15 US-10-027-632-289790 Sequence 289790,
989 16 1.2 580 18 US-10-027-632-50140 Sequence 50140, A
990 16 1.2 583 17 US-10-021-323-2237 Sequence 2237, Ap
991 16 1.2 583 17 US-10-021-323-2237 Sequence 14118, A
992 16 1.2 585 13 US-10-027-632-210856 Sequence 210856,
993 16 1.2 585 13 US-10-027-632-210856 Sequence 210856,
994 16 1.2 585 13 US-10-027-632-210856 Sequence 241067,
995 16 1.2 585 15 US-10-027-632-210856 Sequence 241067,
996 16 1.2 585 15 US-10-027-632-210856 Sequence 241067,
997 16 1.2 586 15 US-10-027-632-211350 Sequence 211350,
998 16 1.2 586 15 US-10-027-632-211350 Sequence 211350,
999 16 1.2 586 13 US-10-027-632-211351 Sequence 211351,
1000 16 1.2 586 13 US-10-027-632-211351 Sequence 211351,
```

ALIGNMENTS

```
RESULT 1
US-10-032-585-6023
; Sequence 6023, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jjiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patent version 3.1
; SEQ ID NO 6023
; LENGTH: 1299
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-6023
```

```
Query Match 68.6%; Score 891; DB 15; Length 1299;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1291; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```

```
Qy 1 ATGTCAAAAGCATTTAGTGCACCTGGAAGCATTTCTTGTGTGATATTTGGTCTT 60
Db 1 ATGTCAAAAGCATTTAGTGCACCTGGAAGCATTTCTTGTGTGATATTTGGTCTT 60
```

```
Qy 61 GAGCCAAATTTATGATGCTTTATGTGACAGCATTTGTCTATCAGAAATGATGCACTTATTAACA 120
|||
Db 61 GAGCCAAATTTATGATGCTTTATGTGACAGCATTTGTCTATCAGAAATGATGCACTTATTAACA 120
121 CCAAAAGGAAACGATTTGAAAGAAATCTAGAAATCAAAATTTCTTCAACCCCAATTTGCAAC 180
|||
Db 121 CCAAAAGGAAACGATTTGAAAGAAATCTAGAAATCAAAATTTCTTCAACCCCAATTTGCAAC 180
122 CCAAAAGGAAACGATTTGAAAGAAATCTAGAAATCAAAATTTCTTCAACCCCAATTTGCAAC 180
|||
Qy 181 GGAAGATGGAAATATCAATATATCAATATCAAGAAAGCCAGAAAGTTCACTGACGC 240
|||
Db 181 GGAAGATGGAAATATCAATATATCAATATCAAGAAAGCCAGAAAGTTCACTGACGC 240
241 ATAAATCCATTTTATGAGCAACTATATATATGCTTATGCTTATATCAACCGACGA 300
|||
Qy 241 ATAAATCCATTTTATGAGCAACTATATATATGCTTATGCTTATATCAACCGACGA 300
|||
Db 241 ATAAATCCATTTTATGAGCAACTATATATATGCTTATGCTTATATCAACCGACGA 300
301 GCATTTGATCTTGAATCATATTTACTCAGACCCCTGATATCATTCACAGAAAGTACT 360
|||
Qy 301 GCATTTGATCTTGAATCATATTTACTCAGACCCCTGATATCATTCACAGAAAGTACT 360
|||
Db 301 GCATTTGATCTTGAATCATATTTACTCAGACCCCTGATATCATTCACAGAAAGTACT 360
361 GAAACCAAGATCTCTGAAATGAGAAAGAAACATTTCTTACATTCCTGTCATTAAC 420
|||
Qy 361 GAAACCAAGATCTCTGAAATGAGAAAGAAACATTTCTTACATTCCTGTCATTAAC 420
|||
Db 361 GAAACCAAGATCTCTGAAATGAGAAAGAAACATTTCTTACATTCCTGTCATTAAC 420
421 GAAAGTGAAGAAAGACCGATAGTTAGTTACGCGCAGAGATTAAGTGTGTCACAAAGT 480
|||
Qy 421 GAAAGTGAAGAAAGACCGATAGTTAGTTACGCGCAGAGATTAAGTGTGTCACAAAGT 480
|||
Db 421 GAAAGTGAAGAAAGACCGATAGTTAGTTACGCGCAGAGATTAAGTGTGTCACAAAGT 480
421 GAAAGTGAAGAAAGACCGATAGTTAGTTACGCGCAGAGATTAAGTGTGTCACAAAGT 480
|||
Qy 421 GAAAGTGAAGAAAGACCGATAGTTAGTTACGCGCAGAGATTAAGTGTGTCACAAAGT 480
|||
Db 421 GAAAGTGAAGAAAGACCGATAGTTAGTTACGCGCAGAGATTAAGTGTGTCACAAAGT 480
481 TTATATCCATTTTATGAGCAACTATATATATGCTTATGCTTATATCAACCGACGA 540
|||
Qy 481 TTATATCCATTTTATGAGCAACTATATATATGCTTATGCTTATATCAACCGACGA 540
|||
Db 481 TTATATCCATTTTATGAGCAACTATATATATGCTTATGCTTATATCAACCGACGA 540
541 GCAAGATGGAACATTTGATATGCAAGAAAGAAAGTATGATGCTTATGATGCTTACT 600
|||
Qy 541 GCAAGATGGAACATTTGATATGCAAGAAAGAAAGTATGATGCTTATGATGCTTACT 600
|||
Db 541 GCAAGATGGAACATTTGATATGCAAGAAAGAAAGTATGATGCTTATGATGCTTACT 600
541 GCAAGATGGAACATTTGATATGCAAGAAAGAAAGTATGATGCTTATGATGCTTACT 600
|||
Qy 541 GCAAGATGGAACATTTGATATGCAAGAAAGAAAGTATGATGCTTATGATGCTTACT 600
|||
Db 541 GCAAGATGGAACATTTGATATGCAAGAAAGAAAGTATGATGCTTATGATGCTTACT 600
601 GCAATTTATGCTTATGATATGATATGATATGATATGATATGATATGATATGATATGAT 660
|||
Qy 601 GCAATTTATGCTTATGATATGATATGATATGATATGATATGATATGATATGATATGAT 660
|||
Db 601 GCAATTTATGCTTATGATATGATATGATATGATATGATATGATATGATATGATATGAT 660
661 CAGGTTCTAGAAAGTATCTGAGAAAGTCCCAAGAGTTGAAAGAAATGATGAAAGT 720
|||
Qy 661 CAGGTTCTAGAAAGTATCTGAGAAAGTCCCAAGAGTTGAAAGAAATGATGAAAGT 720
|||
Db 661 CAGGTTCTAGAAAGTATCTGAGAAAGTCCCAAGAGTTGAAAGAAATGATGAAAGT 720
721 AACTGGGAATTCACAAATGAAAGTATGATATGATATGATATGATATGATATGATATGAT 780
|||
Qy 721 AACTGGGAATTCACAAATGAAAGTATGATATGATATGATATGATATGATATGATATGAT 780
|||
Db 721 AACTGGGAATTCACAAATGAAAGTATGATATGATATGATATGATATGATATGATATGAT 780
721 AACTGGGAATTCACAAATGAAAGTATGATATGATATGATATGATATGATATGATATGAT 780
|||
Qy 721 AACTGGGAATTCACAAATGAAAGTATGATATGATATGATATGATATGATATGATATGAT 780
|||
Db 721 AACTGGGAATTCACAAATGAAAGTATGATATGATATGATATGATATGATATGATATGAT 780
781 GACGTCAAGGTGCTCAGAAACACCCAAATGATATGATATGATATGATATGATATGAT 840
|||
Qy 781 GACGTCAAGGTGCTCAGAAACACCCAAATGATATGATATGATATGATATGATATGAT 840
|||
Db 781 GACGTCAAGGTGCTCAGAAACACCCAAATGATATGATATGATATGATATGATATGAT 840
841 GAAAGCCAAAGAAAGCTGTTGTTGATGATGATGATGATGATGATGATGATGATGAT 900
|||
Qy 841 GAAAGCCAAAGAAAGCTGTTGTTGATGATGATGATGATGATGATGATGATGATGAT 900
|||
Db 841 GAAAGCCAAAGAAAGCTGTTGTTGATGATGATGATGATGATGATGATGATGATGAT 900
901 ATGAAGGAATTTAGGGAATGCGTGAAGAAATAGCACTCAGACCCAGAGACTTATATTA 960
|||
Qy 901 ATGAAGGAATTTAGGGAATGCGTGAAGAAATAGCACTCAGACCCAGAGACTTATATTA 960
|||
Db 901 ATGAAGGAATTTAGGGAATGCGTGAAGAAATAGCACTCAGACCCAGAGACTTATATTA 960
961 GAGTTAGATCATTTCTGTTGAGCTTTGACCTGTTGACATTTAAGAAATCAGAAAGGTT 1020
|||
Qy 961 GAGTTAGATCATTTCTGTTGAGCTTTGACCTGTTGACATTTAAGAAATCAGAAAGGTT 1020
|||
Db 961 GAGTTAGATCATTTCTGTTGAGCTTTGACCTGTTGACATTTAAGAAATCAGAAAGGTT 1020
1021 CAAGCATTTAACAAGAAATCAGAGTTTCAATTTGAACTGATGCTCAAAACCAAGTTGTT 1080
|||
Qy 1021 CAAGCATTTAACAAGAAATCAGAGTTTCAATTTGAACTGATGCTCAAAACCAAGTTGTT 1080
|||
Db 1021 CAAGCATTTAACAAGAAATCAGAGTTTCAATTTGAACTGATGCTCAAAACCAAGTTGTT 1080
1081 GACCGTTTCAAGAGATTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1140
|||
Qy 1081 GACCGTTTCAAGAGATTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1140
|||
Db 1081 GACCGTTTCAAGAGATTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1140
1141 GATGCAATAGCTGATATAGTTGTTGAAATCAAGTGGGAAATTTTAAAGCAAGAACTTT 1200
|||
Qy 1141 GATGCAATAGCTGATATAGTTGTTGAAATCAAGTGGGAAATTTTAAAGCAAGAACTTT 1200
|||
```



```
Db      1141 GATGCAATAGCTATTAGTTGGAATAATCAAGTGGAAATTTTAAGCAGAAACTCTT 1200
Qy      1201 GAAATCCAGATTATTTTATTAATGTTTCTGGGTTGATTGGAGAGCAAGAGGT 1260
Db      1201 GAAATCCAGATTATTTTATTAATGTTTCTGGGTTGATTGGAGAGCAAGAGGT 1260
Qy      1261 GTACTTGAGAAAAAACAGAGACTATATAGTTTATTA 1299
Db      1261 GTACTTGAGAAAAAACAGAGACTATATAGTTTATTA 1299

RESULT 2
US-10-032-585-3023/c
; Sequence 3023, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jjiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patent version 3.1
; SEQ ID NO 3023
; LENGTH: 65
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-3023

Query Match
Best Local Similarity 100.0%; Pred. No. 2e-22; Length 65;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 ATGTCAAAAGCATTTAGTGCACCTGGAAAAAGCATTTCTTGCTGGATATTTGTTCTT 60
Db      65 ATGTCAAAAGCATTTAGTGCACCTGGAAAAAGCATTTCTTGCTGGATATTTGTTCTT 6

Qy      61 GAGCC 65
Db      5 GAGCC 1

RESULT 3
US-10-032-585-5023/c
; Sequence 5023, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jjiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patent version 3.1
; SEQ ID NO 5023
; LENGTH: 22
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-5023

Query Match
Best Local Similarity 100.0%; Pred. No. 2.3; Length 22;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      454 GGATTAGTGTCAAGTTGTTGCCA 475
```

```
Db      22 GGATTAGTGTCAAGTTGTTGCCA 1

RESULT 4
US-10-027-632-150718/c
; Sequence 150718, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150718
; LENGTH: 690
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-150718

Query Match
Best Local Similarity 100.0%; Pred. No. 3.1; Length 690;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1030 ACACAAATTCAGAGGTTCCAA 1051
Db      325 ACACAAATTCAGAGGTTCCAA 304

RESULT 5
US-10-027-632-150718/c
; Sequence 150718, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150718
; LENGTH: 690
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-150718

Query Match
Best Local Similarity 100.0%; Pred. No. 3.1; Length 690;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1030 ACACAAATTCAGAGGTTCCAA 1051
Db      325 ACACAAATTCAGAGGTTCCAA 304

RESULT 5
US-10-027-632-150718/c
; Sequence 150718, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150718
; LENGTH: 690
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-150718

Query Match
Best Local Similarity 100.0%; Pred. No. 3.1; Length 690;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1030 ACACAAATTCAGAGGTTCCAA 1051
Db      325 ACACAAATTCAGAGGTTCCAA 304
```

LENGTH: 690
TYPE: DNA
ORGANISM: Human
US-10-027-632-150718

Query Match
Best Local Similarity 1.7%; Score 22; DB 15; Length 690;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1030 ACACAAAATCAGAGTTCCAA 1051
DB 325 ACACAAAATCAGAGTTCCAA 304

RESULT 6
US-10-027-632-26202/c
Sequence 26202, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIOR FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 26202
LENGTH: 711
TYPE: DNA
ORGANISM: Human
US-10-027-632-26202

Query Match
Best Local Similarity 1.7%; Score 22; DB 13; Length 711;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1030 ACACAAAATCAGAGTTCCAA 1051
DB 325 ACACAAAATCAGAGTTCCAA 304

RESULT 7
US-10-027-632-26203/c
Sequence 26203, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIOR FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 26203
LENGTH: 711
TYPE: DNA
ORGANISM: Human
US-10-027-632-26203

Query Match
Best Local Similarity 1.7%; Score 22; DB 13; Length 711;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1030 ACACAAAATCAGAGTTCCAA 1051
DB 325 ACACAAAATCAGAGTTCCAA 304

RESULT 8
US-10-027-632-26202/c
Sequence 26202, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIOR FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 26202
LENGTH: 711
TYPE: DNA
ORGANISM: Human
US-10-027-632-26202

Query Match
Best Local Similarity 1.7%; Score 22; DB 15; Length 711;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1030 ACACAAAATCAGAGTTCCAA 1051
DB 325 ACACAAAATCAGAGTTCCAA 304

RESULT 9
US-10-027-632-26203/c
Sequence 26203, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.

```
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ TITLE OF INVENTION: Polymorphisms in the Human Genome
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ PRIOR FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 26203
/ LENGTH: 711
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-26203
```

```
Query Match          1.6%; Score 22; DB 15; Length 711;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1030 ACACAAATTCAGAGTCCAA 1051
      |||||||
DB      325 ACACAAATTCAGAGTCCAA 304
```

```
RESULT 10
US-10-242-535A-6452
/ Sequence 6452, Application US/10242535A
/ Publication No. US20040013663A1
/ GENERAL INFORMATION:
/ APPLICANT: ChondroGene Inc.
/ APPLICANT: Liew, C.C.
/ TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
/ FILE REFERENCE: 4231/2005
/ CURRENT APPLICATION NUMBER: US/10/242,535A
/ PRIOR FILING DATE: 2002-09-12
/ PRIOR APPLICATION NUMBER: US 10/085,783
/ PRIOR FILING DATE: 2002-02-28
/ PRIOR APPLICATION NUMBER: US 60/305,340
/ PRIOR FILING DATE: 2001-07-13
/ PRIOR APPLICATION NUMBER: US 60/275,017
/ PRIOR FILING DATE: 2001-03-12
/ PRIOR APPLICATION NUMBER: US 60/271,955
/ PRIOR FILING DATE: 2001-02-28
/ NUMBER OF SEQ ID NOS: 58994
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 6452
/ LENGTH: 313
/ TYPE: DNA
/ ORGANISM: Human
US-10-242-535A-6452
```

```
Query Match          1.6%; Score 21; DB 16; Length 313;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1115 GTGTGTTCCAGTGTGTG 1135
      |||||||
DB      36 GTGTGTTCCAGTGTGTG 56
```

RESULT 11

```
US-10-085-783A-6452
/ Sequence 6452, Application US/10085783A
/ Publication No. US20040037841A1
/ GENERAL INFORMATION:
/ APPLICANT: ChondroGene Inc.
/ APPLICANT: Liew, C.C.
/ TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
/ FILE REFERENCE: 4231/2002
/ CURRENT APPLICATION NUMBER: US/10/085,783A
/ PRIOR FILING DATE: 2002-02-28
/ PRIOR APPLICATION NUMBER: US 60/305,340
/ PRIOR FILING DATE: 2001-07-13
/ PRIOR APPLICATION NUMBER: US 60/275,017
/ PRIOR FILING DATE: 2001-03-12
/ PRIOR APPLICATION NUMBER: US 60/271,955
/ PRIOR FILING DATE: 2001-02-28
/ NUMBER OF SEQ ID NOS: 58994
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 6452
/ LENGTH: 313
/ TYPE: DNA
/ ORGANISM: Human
US-10-085-783A-6452
```

```
Query Match          1.6%; Score 21; DB 16; Length 313;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1115 GTGTGTTCCAGTGTGTG 1135
      |||||||
DB      36 GTGTGTTCCAGTGTGTG 56
```

```
RESULT 12
US-10-767-795-5310
/ Sequence 5310, Application US/10767795
/ Publication No. US20040181830A1
/ GENERAL INFORMATION:
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Zhou, Yihua
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
/ FILE REFERENCE: 38-21(53534)B
/ CURRENT APPLICATION NUMBER: US/10/767,795
/ PRIOR FILING DATE: 2004-01-30
/ NUMBER OF SEQ ID NOS: 117596
/ SEQ ID NO 5310
/ LENGTH: 400
/ TYPE: DNA
/ ORGANISM: Gossypium hirsutum
/ FEATURE:
/ OTHER INFORMATION: Clone ID: GOSHI-09MAY01-C61781_1
US-10-767-795-5310
```

```
Query Match          1.6%; Score 21; DB 17; Length 400;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1115 GTGTGTTCCAGTGTGTG 1135
      |||||||
DB      238 GTGTGTTCCAGTGTGTG 258
```

```
RESULT 13
US-09-918-995-17749
/ Sequence 17749, Application US/09918995
/ Publication No. US20030073623A1
/ GENERAL INFORMATION:
/ APPLICANT: HySeq, Inc.
/ TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
/ TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
/ FILE REFERENCE: 20411-756
```

```

; CURRENT APPLICATION NUMBER: US/09/918,995
; TYPE: DNA
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17749
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1) ... (476)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-17749
```

```

Query Match
Best Local Similarity 100.0%; Score 21; DB 10; Length 476;
Pred. No. 9,7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1115 GTGTGTTCCAGTGTCTGTG 1135
DB 126 GTGTGTTCCAGTGTCTGTG 146
```

```

RESULT 14
US-09-920-300A-726
; Sequence 726, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jianshun
; APPLICANT: Secret, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 726
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-726
```

```

Query Match
Best Local Similarity 100.0%; Score 21; DB 9; Length 479;
Pred. No. 9,7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1115 GTGTGTTCCAGTGTCTGTG 1135
DB 41 GTGTGTTCCAGTGTCTGTG 61
```

```

RESULT 15
US-10-033-528-726
; Sequence 726, Application US/10033528
; Publication No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jianshun
; APPLICANT: Secret, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033,528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 726
```

```

; LENGTH: 479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-726
```

```

Query Match
Best Local Similarity 100.0%; Score 21; DB 13; Length 479;
Pred. No. 9,7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1115 GTGTGTTCCAGTGTCTGTG 1135
DB 41 GTGTGTTCCAGTGTCTGTG 61
```

```

RESULT 16
US-10-099-926-726
; Sequence 726, Application US/10099926
; Publication No. US2003016064A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jianshun
; APPLICANT: Secret, Heather
; APPLICANT: Jians, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547C2
; CURRENT APPLICATION NUMBER: US/10/099,926
; CURRENT FILING DATE: 2002-03-17
; NUMBER OF SEQ ID NOS: 1982
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 726
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-099-926-726
```

```

Query Match
Best Local Similarity 100.0%; Score 21; DB 15; Length 479;
Pred. No. 9,7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1115 GTGTGTTCCAGTGTCTGTG 1135
DB 41 GTGTGTTCCAGTGTCTGTG 61
```

```

RESULT 17
US-09-998-598-263/C
; Sequence 263, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jianshun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 263
; LENGTH: 518
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 23, 46, 60, 64, 83, 482
; OTHER INFORMATION: n = A,T,C or G
US-09-998-598-263
```

```

Query Match
1.6%; Score 21; DB 9; Length 518;
```

Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1115 GTGTGTTCCAGTGTCTGTG 1135
Db 414 GTGTGTTCCAGTGTCTGTG 394

RESULT 18
US-09-920-300A-1254/c
; Sequence 1254, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1254
; LENGTH: 530
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-1254

Query Match 1.6%; Score 21; DB 9; Length 530;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1115 GTGTGTTCCAGTGTCTGTG 1135
Db 439 GTGTGTTCCAGTGTCTGTG 419

RESULT 19
US-10-033-528-1254/c
; Sequence 1254, Application US/10033528
; Publication No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033,528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1254
; LENGTH: 530
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-1254

Query Match 1.6%; Score 21; DB 13; Length 530;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1115 GTGTGTTCCAGTGTCTGTG 1135
Db 439 GTGTGTTCCAGTGTCTGTG 419

RESULT 20
US-10-099-926-1254/c
; Sequence 1254, Application US/10099926

; Publication No. US20030166064A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; APPLICANT: Jiang, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547C2
; CURRENT APPLICATION NUMBER: US/10/099,926
; CURRENT FILING DATE: 2002-03-17
; NUMBER OF SEQ ID NOS: 1982
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1254
; LENGTH: 530
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-099-926-1254

Query Match 1.6%; Score 21; DB 15; Length 530;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1115 GTGTGTTCCAGTGTCTGTG 1135
Db 439 GTGTGTTCCAGTGTCTGTG 419

RESULT 21
US-10-198-846-10192
; Sequence 10192, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10192
; LENGTH: 1060
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-10192

Query Match 1.6%; Score 21; DB 14; Length 1060;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1115 GTGTGTTCCAGTGTCTGTG 1135
Db 477 GTGTGTTCCAGTGTCTGTG 497

RESULT 22
US-10-264-049-594
; Sequence 594, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04

```

; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 594
; LENGTH: 1574
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1467)..(1467)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1475)..(1475)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1485)..(1485)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1520)..(1520)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1550)..(1550)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1572)..(1574)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-264-049-594

Query Match
Best Local Similarity 1.6%; Score 21; DB 16; Length 1574;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTGTGTGTG 1135
DB 701 GTGTGTTCCAGTGTGTGTG 721

RESULT 23
US-10-723-860-5689
; Sequence 5689, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Nacasha
; APPLICANT: Ginsburg, Wendy M.
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882,0193.NUS01
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5689
; LENGTH: 1825
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-5689

Query Match
Best Local Similarity 1.6%; Score 21; DB 18; Length 1825;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTGTGTGTG 1135
```

```

DB 448 GTGTGTTCCAGTGTGTGTG 468

RESULT 24
US-10-170-385-414
; Sequence 414, Application US/10170385
; Publication No. US20030203372A1
; GENERAL INFORMATION:
; APPLICANT: Ward, Neil Raymond
; APPLICANT: Mundy, Christopher Robert
; APPLICANT: Kan, On
; APPLICANT: Harris, Robert Alan
; APPLICANT: White, Jonathan
; APPLICANT: Binley, Katie Mary
; APPLICANT: Rayner, William Nigel
; APPLICANT: Naylor, Stuart
; APPLICANT: Kingsman, Susan Mary
; APPLICANT: Krige, David
; TITLE OF INVENTION: ANALYSIS METHOD
; FILE REFERENCE: 53682000100
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: PCT/GB02/01662
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: PCT/GB01/05458
; NUMBER OF SEQ ID NOS: 549
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 414
; LENGTH: 2010
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-170-385-414

Query Match
Best Local Similarity 1.6%; Score 21; DB 15; Length 2010;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTGTGTGTG 1135
DB 1303 GTGTGTTCCAGTGTGTGTG 1323

RESULT 25
US-10-172-118-662
; Sequence 662, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; NUMBER OF SEQ ID NOS: 2659
; SEQ ID NO 662
; LENGTH: 2010
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_001762
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-662
```

Query Match 1.6%; Score 21; DB 15; Length 2010;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTGTCTGTG 1135
|||||
DB 1303 GTGTGTTCCAGTGTCTGTG 1323

RESULT 26

US-10-159-563-393
; Sequence 393, Application US/10159563
; Publication No. US20040009154A1
; GENERAL INFORMATION:
; APPLICANT: Khan, Javed
; APPLICANT: Ringner, Markus
; APPLICANT: Peterson, Carsten
; APPLICANT: Meltzer, Paul
; TITLE OF INVENTION: SELECTIONS OF GENES AND METHODS OF USING THE SAME FOR
; TITLE OF INVENTION: DIAGNOSIS AND FOR TARGETING THE THERAPY OF SELECT CANCERS
; FILE REFERENCE: 11613.56US11
; CURRENT APPLICATION NUMBER: US/10/159,563
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US 10/133,937
; PRIOR FILING DATE: 2002-04-25
; NUMBER OF SEQ ID NOS: 444
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 393
; LENGTH: 2010
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-159-563-393

Query Match 1.6%; Score 21; DB 16; Length 2010;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTGTCTGTG 1135
|||||
DB 1303 GTGTGTTCCAGTGTCTGTG 1323

RESULT 27

US-10-342-887-662
; Sequence 662, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO: 662
; LENGTH: 2010
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-662

Query Match 1.6%; Score 21; DB 16; Length 2010;

Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTGTCTGTG 1135
|||||
DB 1303 GTGTGTTCCAGTGTCTGTG 1323

RESULT 28

US-09-925-300-276
; Sequence 276, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Nucleic acids, proteins and antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 276
; LENGTH: 2594
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2198)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-276

Query Match 1.6%; Score 21; DB 9; Length 2594;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTGTCTGTG 1135
|||||
DB 1298 GTGTGTTCCAGTGTCTGTG 1318

RESULT 29

US-10-357-930-24803
; Sequence 24803, Application US/10357930
; Publication No. US20040259086A1
; GENERAL INFORMATION:
; APPLICANT: Endegge, Robert
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; FILE REFERENCE: MRI-007BCN
; CURRENT APPLICATION NUMBER: US/10/357,930
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: 09/785,276
; PRIOR FILING DATE: 2003-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232


```
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 24803
/ LENGTH: 2837
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-357-930-24803

Query Match      1.6%; Score 21; DB 18; Length 2837;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1115 GTGTGCTCAGGTGCTGATG 1135
Db      1307 GTGTGCTCAGGTGCTGATG 1327

RESULT 30
US-09-796-692-9039
/ Sequence 9039, Application US/09796692
/ Publication No. US20020198362A1
/ GENERAL INFORMATION:
/ APPLICANT: Gaiger, Alexander
/ APPLICANT: Algate, Paul A.
/ APPLICANT: Mannion, Jane
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
/ FILE REFERENCE: 2077.001200
/ CURRENT APPLICATION NUMBER: US/09/796,692
/ PRIOR FILING DATE: 2001-03-01
/ PRIOR APPLICATION NUMBER: 60/186,126
/ PRIOR FILING DATE: 2000-03-01
/ PRIOR APPLICATION NUMBER: 60/190,479
/ PRIOR FILING DATE: 2000-03-17
/ PRIOR APPLICATION NUMBER: 60/200,545
/ PRIOR FILING DATE: 2000-04-27
/ PRIOR APPLICATION NUMBER: 60/200,303
/ PRIOR FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: 60/200,779
/ PRIOR FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: 60/200,999
/ PRIOR FILING DATE: 2000-05-01
/ PRIOR APPLICATION NUMBER: 60/202,084
/ PRIOR FILING DATE: 2000-05-04
/ PRIOR APPLICATION NUMBER: 60/206,201
/ PRIOR FILING DATE: 2000-05-22
/ PRIOR APPLICATION NUMBER: 60/218,950
/ PRIOR FILING DATE: 2000-07-14
/ PRIOR APPLICATION NUMBER: 60/222,903
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: 60/223,416
/ PRIOR FILING DATE: 2000-08-04
/ PRIOR APPLICATION NUMBER: 60/223,378
/ PRIOR FILING DATE: 2000-08-07
/ NUMBER OF SEQ ID NOS: 9597
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 9039
/ LENGTH: 371
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-796-692-9039

Query Match      1.5%; Score 20; DB 9; Length 371;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
/ Publication No. US20030078396A1
/ GENERAL INFORMATION:
/ APPLICANT: Gaiger, Alexander
/ APPLICANT: Algate, Paul A.
/ APPLICANT: Mannion, Jane
/ APPLICANT: Ketter, Marc
/ APPLICANT: Corixa Corporation
/ TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
/ FILE REFERENCE: 014058-013520US
/ CURRENT APPLICATION NUMBER: US/10/040,862
/ PRIOR FILING DATE: 2001-11-06
/ PRIOR APPLICATION NUMBER: US 60/186,126
/ PRIOR FILING DATE: 2000-03-01
/ PRIOR APPLICATION NUMBER: US 60/190,479
/ PRIOR FILING DATE: 2000-03-17
/ PRIOR APPLICATION NUMBER: US 60/200,545
/ PRIOR FILING DATE: 2000-04-27
/ PRIOR APPLICATION NUMBER: US 60/200,303
/ PRIOR FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: US 60/200,779
/ PRIOR FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: US 60/200,999
/ PRIOR FILING DATE: 2000-05-01
/ PRIOR APPLICATION NUMBER: US 60/202,084
/ PRIOR FILING DATE: 2000-05-04
/ PRIOR APPLICATION NUMBER: US 60/206,201
/ PRIOR FILING DATE: 2000-05-22
/ PRIOR APPLICATION NUMBER: US 60/218,950
/ PRIOR FILING DATE: 2000-07-14
/ PRIOR APPLICATION NUMBER: US 60/222,903
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: US 60/223,416
/ PRIOR FILING DATE: 2000-08-04
/ PRIOR APPLICATION NUMBER: US 60/223,378
/ PRIOR FILING DATE: 2000-08-07
/ PRIOR APPLICATION NUMBER: US 09/796,692
/ PRIOR FILING DATE: 2001-03-01
/ NUMBER OF SEQ ID NOS: 10467
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 9039
/ LENGTH: 371
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-040-862-9039

Query Match      1.5%; Score 20; DB 14; Length 371;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      723 CTGGGAATTCAACATGAAA 742
Db      139 CTGGGAATTCAACATGAAA 158

RESULT 32
US-10-057-475B-9039
/ Sequence 9039, Application US/10057475B
/ Publication No. US20040002068A1
/ GENERAL INFORMATION:
/ APPLICANT: Gaiger, Alexander
/ APPLICANT: Algate, Paul A.
/ APPLICANT: Mannion, Jane
/ APPLICANT: Clapper, Jonathan David
/ APPLICANT: Wang, Aijun
/ APPLICANT: Ordenez, Nadia
/ APPLICANT: Carter, Lauren
/ APPLICANT: McNeill, Patricia Dianne
/ APPLICANT: Corixa Corporation
/ TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
/ FILE REFERENCE: 014058-014402US
/ CURRENT APPLICATION NUMBER: US/10/057,475B
```

```
/ CURRENT FILING DATE: 2002-01-22
/ PRIOR APPLICATION NUMBER: US 60/186,126
/ PRIOR FILING DATE: 2000-03-01
/ PRIOR APPLICATION NUMBER: US 60/190,479
/ PRIOR FILING DATE: 2000-03-17
/ PRIOR APPLICATION NUMBER: US 60/200,545
/ PRIOR FILING DATE: 2000-04-27
/ PRIOR APPLICATION NUMBER: US 60/200,303
/ PRIOR FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: US 60/200,779
/ PRIOR FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: US 60/200,999
/ PRIOR FILING DATE: 2000-05-01
/ PRIOR APPLICATION NUMBER: US 60/202,084
/ PRIOR FILING DATE: 2000-05-04
/ PRIOR APPLICATION NUMBER: US 60/206,201
/ PRIOR FILING DATE: 2000-05-22
/ PRIOR APPLICATION NUMBER: US 60/218,950
/ PRIOR FILING DATE: 2000-07-14
/ PRIOR APPLICATION NUMBER: US 60/222,903
/ PRIOR FILING DATE: 2000-08-03
/ Remaining Prior Application data removed - See file wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 10979
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO: 9039
/ LENGTH: 371
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-057-475B-9039
```

```
Query Match          1.5%; Score 20; DB 16; Length 371;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 723 CTGGGAATTCAAACATGAAA 742
|||||
Db 139 CTGGGAATTCAAACATGAAA 158
```

```
RESULT 33
US-10-154-884B-9039
/ Sequence 9039, Application US/10154884B
/ Publication No. US20040005561A1
/ GENERAL INFORMATION:
/ APPLICANT: Galiger, Alexander
/ APPLICANT: Algate, Paul A.
/ APPLICANT: Mannion, Jane
/ APPLICANT: Retter, Marc W.
/ APPLICANT: Corixa Corporation
/ TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
/ FILE REFERENCE: 014058-013521US
/ CURRENT APPLICATION NUMBER: US/10/154,884B
/ PRIOR FILING DATE: 2002-05-23
/ PRIOR APPLICATION NUMBER: US 60/186,126
/ PRIOR FILING DATE: 2000-03-01
/ PRIOR APPLICATION NUMBER: US 60/190,479
/ PRIOR FILING DATE: 2000-03-17
/ PRIOR APPLICATION NUMBER: US 60/200,545
/ PRIOR FILING DATE: 2000-04-27
/ PRIOR APPLICATION NUMBER: US 60/200,303
/ PRIOR FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: US 60/200,779
/ PRIOR FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: US 60/200,999
/ PRIOR FILING DATE: 2000-05-01
/ PRIOR APPLICATION NUMBER: US 60/202,084
/ PRIOR FILING DATE: 2000-05-04
/ PRIOR APPLICATION NUMBER: US 60/206,201
/ PRIOR FILING DATE: 2000-05-22
/ PRIOR APPLICATION NUMBER: US 60/218,950
/ PRIOR FILING DATE: 2000-07-14
/ PRIOR APPLICATION NUMBER: US 60/222,903
```

```
/ PRIOR FILING DATE: 2000-08-03
/ Remaining Prior Application data removed - See file wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 11290
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO: 9039
/ LENGTH: 371
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-154-884B-9039
```

```
Query Match          1.5%; Score 20; DB 16; Length 371;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 723 CTGGGAATTCAAACATGAAA 742
|||||
Db 139 CTGGGAATTCAAACATGAAA 158
```

```
RESULT 34
US-10-764-324-9039
/ Sequence 9039, Application US/10764324
/ Publication No. US20040175739A1
/ GENERAL INFORMATION:
/ APPLICANT: Galiger, Alexander
/ APPLICANT: Algate, Paul A.
/ APPLICANT: Mannion, Jane
/ APPLICANT: Retter, Marc
/ APPLICANT: Corixa Corporation
/ TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
/ FILE REFERENCE: 014058-013520US
/ CURRENT APPLICATION NUMBER: US/10/764,324
/ PRIOR FILING DATE: 2004-01-23
/ PRIOR APPLICATION NUMBER: US/10/040,862
/ PRIOR FILING DATE: 2001-11-06
/ PRIOR APPLICATION NUMBER: US 60/186,126
/ PRIOR FILING DATE: 2000-03-01
/ PRIOR APPLICATION NUMBER: US 60/190,479
/ PRIOR FILING DATE: 2000-03-17
/ PRIOR APPLICATION NUMBER: US 60/200,545
/ PRIOR FILING DATE: 2000-04-27
/ PRIOR APPLICATION NUMBER: US 60/200,303
/ PRIOR FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: US 60/200,779
/ PRIOR FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: US 60/200,999
/ PRIOR FILING DATE: 2000-05-01
/ PRIOR APPLICATION NUMBER: US 60/202,084
/ PRIOR FILING DATE: 2000-05-04
/ PRIOR APPLICATION NUMBER: US 60/206,201
/ PRIOR FILING DATE: 2000-05-22
/ PRIOR APPLICATION NUMBER: US 60/218,950
/ PRIOR FILING DATE: 2000-07-14
/ Remaining Prior Application data removed - See file wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 10467
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO: 9039
/ LENGTH: 371
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-764-324-9039
```

```
Query Match          1.5%; Score 20; DB 17; Length 371;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 723 CTGGGAATTCAAACATGAAA 742
|||||
Db 139 CTGGGAATTCAAACATGAAA 158
```

RESULT 35

```
US-10-029-386-3406/c
; Sequence 3406, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEWICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; NUMBER OF SEQ ID NOS: 34286
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 3406
; LENGTH: 502
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC015845.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
; OTHER INFORMATION: SWISSPROT HIT: Q9Y3D9, EVALUE 2.00e-29
; OTHER INFORMATION: NT HIT: AF161447.1, EVALUE 2.00e-92
; US-10-029-386-3406
```

```
Query Match 1.5%; Score 20; DB 15; Length 502;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 723 CTGGGATTCACACATGAAA 742
Db 95 CTGGGATTCACACATGAAA 76
```

```
RESULT 36
US-09-938-842A-601
; Sequence 601, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 601
; LENGTH: 630
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-601
```

```
Query Match 1.5%; Score 20; DB 9; Length 630;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 41 CTGGTGATATTTGGTTCTT 60
|||||
```

```
Db 317 CTGGTGATATTTGGTTCTT 336
```

```
RESULT 37
US-09-938-842A-601
; Sequence 601, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 601
; LENGTH: 630
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-601
```

```
Query Match 1.5%; Score 20; DB 11; Length 630;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 41 CTGGTGATATTTGGTTCTT 60
Db 317 CTGGTGATATTTGGTTCTT 336
```

```
RESULT 38
US-10-027-632-160412/c
; Sequence 160412, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108627.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 160412
; LENGTH: 728
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-160412
```

```
Query Match 1.5%; Score 20; DB 13; Length 728;
```

Best Local Similarity 100.0%; Pred. No. 33;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1117 GTGCTTCAGGCTGCTGTGG 1136
 |||||
 DB 466 GTGCTTCAGGCTGCTGTGG 447

RESULT 39
 US-10-027-632-160412/c
 ; Sequence 160412, Application US/10027632
 ; Publication No. US20030204075A9
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, David G.
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
 ; TITLE OF INVENTION: Polymorphisms in the Human Genome
 ; FILE REFERENCE: 108827.129
 ; CURRENT APPLICATION NUMBER: US/10/027,632
 ; CURRENT FILING DATE: 2002-04-30
 ; PRIOR APPLICATION NUMBER: US 60/218,006
 ; PRIOR FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: US 60/198,676
 ; PRIOR FILING DATE: 2000-04-20
 ; PRIOR APPLICATION NUMBER: US 60/193,483
 ; PRIOR FILING DATE: 2000-03-29
 ; PRIOR APPLICATION NUMBER: US 60/185,218
 ; PRIOR FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/167,363
 ; PRIOR FILING DATE: 1998-11-23
 ; PRIOR APPLICATION NUMBER: US 60/156,358
 ; PRIOR FILING DATE: 1999-09-28
 ; PRIOR APPLICATION NUMBER: US 60/146,002
 ; PRIOR FILING DATE: 1999-08-09
 ; NUMBER OF SEQ ID NOS: 325720
 ; SOFTWARE: FASTSEQ for Windows Version 4.0
 ; SEQ ID NO 160412
 ; LENGTH: 728
 ; TYPE: DNA
 ; ORGANISM: Human
 US-10-027-632-160412

Query Match 1.5%; Score 20; DB 15; Length 728;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1117 GTGCTTCAGGCTGCTGTGG 1136
 |||||
 DB 466 GTGCTTCAGGCTGCTGTGG 447

RESULT 40
 US-10-369-493-43499
 ; Sequence 43499, Application US/10369493
 ; Publication No. US2003023675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianteng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 43499
 ; LENGTH: 1656
 ; TYPE: DNA
 ; ORGANISM: No. US2003023675A1loc punctiforme
 US-10-369-493-43499

Query Match 1.5%; Score 20; DB 15; Length 1656;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 697 GAGTTGAAAAAATTGATTGA 716
 |||||
 DB 571 GAGTTGAAAAAATTGATTGA 590

Search completed: January 26, 2005, 11:17:55
 Job time : 602 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 26, 2005, 07:43:20 ; Search time 3060 Seconds
(without alignments)
15469.023 Million cell updates/sec

Title: US-10-069-062-6

Perfect score: 1239

Sequence: 1 atgcacaaagcattcgtgc.....aagactataggtttacaa 1299

Scoring table: OLIGO_NUC

Searched: 32822875 seqs, 18219865908 residues

Word size: 0

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database: EST:

1: gb_esc1:*
2: gb_esc2:*
3: gb_esc3:*
4: gb_esc4:*
5: gb_esc5:*
6: gb_esc6:*
7: gb_esc7:*
8: gb_esc8:*
9: gb_esc9:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26	2.0	730	4	BG470614 602511536
2	23	1.8	709	2	AM193927 xm19h03.x
3	23	1.8	1454	9	CL641530 CH213-13H
4	22	1.7	542	8	AQ663605 HS-3476-A
5	21	1.6	531	4	BM705592 UI-E-DWO-
6	21	1.6	349	1	AV684301 AV684301
7	21	1.6	349	1	AV697455 AV697455
8	21	1.6	349	1	AV697455 AV697455
9	21	1.6	349	1	AV697455 AV697455
10	21	1.6	368	4	BI036922 MR4-NT014
11	21	1.6	368	4	BM762828 K-EST0044
12	21	1.6	368	4	BM762828 K-EST0044
13	21	1.6	371	5	BP623179 BP623179
14	21	1.6	372	2	AW361182 RCI-CT025
15	21	1.6	386	6	CF141588 UI-HF-CR0
16	21	1.6	414	7	CN345987 170006000
17	21	1.6	424	7	AA150852 z144b05.x
18	21	1.6	425	7	W69312 z44b05.x
19	21	1.6	426	5	BO347968 CM3-HT019
20	21	1.6	426	7	T66047 YC79e07.x1
21	21	1.6	434	4	BM853177 K-EST0134
22	21	1.6	434	6	CB117286 K-EST0162
23	21	1.6	441	1	AI910561 W68607.x
24	21	1.6	443	1	AA127935 z114E07.x

25	21	1.6	459	4	BG761673 602718849
26	21	1.6	493	2	BE392663 601307517
27	21	1.6	499	4	BM841588 K-EST0118
28	21	1.6	504	2	BE265418 601194958
29	21	1.6	506	7	CN345984 170006001
30	21	1.6	511	6	CB159634 K-EST0219
31	21	1.6	516	4	BM796788 K-EST0079
32	21	1.6	522	4	BG149625 nad31d12.
33	21	1.6	528	6	CB124236 K-EST0172
34	21	1.6	536	7	W38880 zb22a05.x1
35	21	1.6	542	4	BM840031 K-EST0117
36	21	1.6	543	6	CB119231 K-EST0166
37	21	1.6	547	6	CB147674 K-EST0203
38	21	1.6	548	1	AA186456 zp70h10.x
39	21	1.6	553	6	CB160975 K-EST0220
40	21	1.6	553	8	AO837961 HS_4647.B
41	21	1.6	556	4	BM279286 NB_ad1_05
42	21	1.6	561	2	BE729211 601560832
43	21	1.6	566	6	CB113018 K-EST0154
44	21	1.6	568	2	BE543072 601069008
45	21	1.6	570	6	CB160702 K-EST0220
46	21	1.6	576	1	AV693973 AV693973
47	21	1.6	576	9	CE64965 c1gr-g88-
48	21	1.6	579	4	BM840345 K-EST0117
49	21	1.6	579	5	BQ348328 IL0-HT020
50	21	1.6	585	2	BF576343 602134224
51	21	1.6	589	4	BM841432 K-EST0118
52	21	1.6	596	9	AG226531 Lotus cor
53	21	1.6	597	6	CB270026 1008933.H
54	21	1.6	603	7	W68256 z235d05.x1
55	21	1.6	606	5	BQ348329 IL0-HT020
56	21	1.6	607	6	CA341142 haa79b03.
57	21	1.6	609	2	AM369207 IL0-HT020
58	21	1.6	609	5	BY460801 BY460801
59	21	1.6	611	6	CB124509 K-EST0173
60	21	1.6	617	4	BM841617 K-EST0118
61	21	1.6	623	6	CD678902 hq01f12.y
62	21	1.6	625	2	BF237622 601841980
63	21	1.6	629	7	CN345974 170006001
64	21	1.6	630	6	CD676885 hoo07h03.y
65	21	1.6	632	4	BM794782 K-EST0076
66	21	1.6	632	6	CB110558 K-EST0152
67	21	1.6	632	7	CN345973 170004554
68	21	1.6	632	7	CN345973 170005336
69	21	1.6	635	2	BG762324 602737226
70	21	1.6	638	5	BQ347584 CM4-HT019
71	21	1.6	644	2	AM411390 fh12a10.x
72	21	1.6	648	6	CA418220 UI-H-FHO-
73	21	1.6	650	2	BE615514 601278835
74	21	1.6	651	2	BE535183 601058709
75	21	1.6	655	7	BM478815 UI-CF-FNO
76	21	1.6	658	4	BM781898 K-EST0058
77	21	1.6	663	6	CA418049 UI-H-FHO-
78	21	1.6	668	6	CA419487 UI-H-FHO-
79	21	1.6	680	4	BI850890 imageqc.1
80	21	1.6	680	4	BI850890 imageqc.1
81	21	1.6	691	1	AU116915 AU116915
82	21	1.6	691	2	BE390849 601286566
83	21	1.6	692	5	BQ009693 UI-H-ED1-
84	21	1.6	693	4	BI223761 602943145
85	21	1.6	705	2	BE251866 60115895
86	21	1.6	708	6	AU119388 AU119388
87	21	1.6	708	6	CF121286 UI-HF-CR0
88	21	1.6	714	7	CN345956 170004708
89	21	1.6	719	6	CA439945 UI-H-D10-
90	21	1.6	722	9	CN418039 170004245
91	21	1.6	722	9	AG357605 Mus muscu
92	21	1.6	727	1	AU137581 AU137581
93	21	1.6	727	4	BG830296 602764692
94	21	1.6	731	5	BO774219 UI-H-FHO-
95	21	1.6	733	2	BE563941 601348178
96	21	1.6	735	4	BG475173 602491736
97	21	1.6	742	1	AU126169 AU126169

C 390	19	1.5	763	9	AG440581	C 463	18	1.4	183	7	CR284785
C 391	19	1.5	764	5	BU333121	464	18	1.4	183	7	CR567662
C 392	19	1.5	765	7	CO061585	465	18	1.4	194	2	BF475477
C 393	19	1.5	774	4	BI116688	466	18	1.4	194	2	BF475477
C 394	19	1.5	776	5	BO295188	467	18	1.4	202	1	AV005929
C 395	19	1.5	788	6	CA346021	468	18	1.4	206	1	AI382688
C 396	19	1.5	789	5	CG697840	469	18	1.4	209	9	CR478255
C 397	19	1.5	798	5	BM859214	470	18	1.4	213	5	BM327484
C 398	19	1.5	809	6	CF148664	471	18	1.4	216	6	CD195070
C 399	19	1.5	811	6	CA368558	472	18	1.4	225	1	AV346678
C 400	19	1.5	821	8	BZ01501	473	18	1.4	231	8	AE742304
C 401	19	1.5	825	8	BZ97872	474	18	1.4	240	7	DE0887
C 402	19	1.5	826	5	BX15978	475	18	1.4	252	8	AE058004
C 403	19	1.5	827	8	AO745023	476	18	1.4	253	4	BI965935
C 404	19	1.5	828	4	BI103926	477	18	1.4	254	1	AV082240
C 405	19	1.5	834	5	BU226259	478	18	1.4	258	5	BO843353
C 406	19	1.5	835	7	CK196698	479	18	1.4	260	4	BO037951
C 407	19	1.5	837	7	CG521409	480	18	1.4	266	2	BE496840
C 408	19	1.5	840	9	CC720356	481	18	1.4	266	7	CO539318
C 409	19	1.5	842	6	CA119655	482	18	1.4	267	2	BB608655
C 410	19	1.5	842	6	CG949573	483	18	1.4	269	2	BB343492
C 411	19	1.5	843	8	BM592722	484	18	1.4	271	5	BM117941
C 412	19	1.5	849	5	BU291337	485	18	1.4	274	7	CR683578
C 413	19	1.5	858	7	CR291904	486	18	1.4	275	7	CR683578
C 414	19	1.5	860	9	CG969483	487	18	1.4	279	1	AV139640
C 415	19	1.5	862	2	BF262461	488	18	1.4	280	5	BM248630
C 416	19	1.5	882	2	BF262461	489	18	1.4	281	2	BF192670
C 417	19	1.5	883	9	CG448087	490	18	1.4	283	4	BG017975
C 418	19	1.5	893	9	CC706200	491	18	1.4	288	4	CG060275
C 419	19	1.5	895	9	CG720367	492	18	1.4	298	4	CG1653275
C 420	19	1.5	896	6	AG141027	493	18	1.4	301	9	CD1653275
C 421	19	1.5	904	8	BM164773	494	18	1.4	308	6	CD189171
C 422	19	1.5	904	8	BM164773	495	18	1.4	309	6	CD189171
C 423	19	1.5	914	5	EX433074	496	18	1.4	312	7	F37104
C 424	19	1.5	916	9	CG092484	497	18	1.4	313	8	BZ215895
C 425	19	1.5	918	9	CG988557	498	18	1.4	316	1	AA655406
C 426	19	1.5	920	9	CNS01PMA	499	18	1.4	317	2	BG426418
C 427	19	1.5	924	9	CG67005	500	18	1.4	319	7	BM243905
C 428	19	1.5	926	8	BM155622	501	18	1.4	322	9	TE9270
C 429	19	1.5	929	5	BQ650256	502	18	1.4	324	2	CL579708
C 430	19	1.5	929	9	CG388749	503	18	1.4	325	2	BE242221
C 431	19	1.5	937	9	CG396006	504	18	1.4	328	7	F27990
C 432	19	1.5	940	8	AZ674704	505	18	1.4	336	9	CR146580
C 433	19	1.5	942	8	CC441740	506	18	1.4	344	7	WM4413
C 434	19	1.5	950	7	CF412980	507	18	1.4	344	9	PCB303914
C 435	19	1.5	957	9	CG367014	508	18	1.4	345	1	AA902540
C 436	19	1.5	967	9	CG382383	509	18	1.4	348	7	AE144185
C 437	19	1.5	969	5	CG184068	510	18	1.4	349	4	BM133623
C 438	19	1.5	980	5	BU237232	511	18	1.4	354	1	AA097944
C 439	19	1.5	987	8	CC012894	512	18	1.4	354	4	BM692922
C 440	19	1.5	1079	8	CC256850	513	18	1.4	362	6	CE68114
C 441	19	1.5	1096	9	CNS056CJ	514	18	1.4	367	4	AI169131
C 442	19	1.5	1101	9	CNS051SN	515	18	1.4	367	4	BM692922
C 443	19	1.5	1105	8	CC217038	516	18	1.4	368	1	AA159444
C 444	19	1.5	1111	9	CL043555	517	18	1.4	370	6	CD714551
C 445	19	1.5	1188	9	CL118724	518	18	1.4	370	6	CD714551
C 446	19	1.5	1189	1	AL518120	519	18	1.4	370	6	CD714551
C 447	19	1.5	1206	8	CC266664	520	18	1.4	370	6	CD714551
C 448	19	1.5	1229	9	CL051231	521	18	1.4	374	6	CA653223
C 449	19	1.5	1260	9	AG127510	522	18	1.4	376	6	CA653223
C 450	19	1.5	1269	9	AG127510	523	18	1.4	376	6	CA653223
C 451	19	1.5	1459	9	CG754198	524	18	1.4	376	6	CA653223
C 452	19	1.5	1657	2	BF128477	525	18	1.4	376	6	CA653223
C 453	19	1.5	1857	9	CG756776	526	18	1.4	376	6	CA653223
C 454	19	1.5	1872	9	CG754248	527	18	1.4	376	6	CA653223
C 455	19	1.5	1872	9	CG754248	528	18	1.4	376	6	CA653223
C 456	19	1.5	1872	9	CG754248	529	18	1.4	376	6	CA653223
C 457	19	1.5	1872	9	CG754248	530	18	1.4	376	6	CA653223
C 458	19	1.5	1872	9	CG754248	531	18	1.4	376	6	CA653223
C 459	19	1.5	1872	9	CG754248	532	18	1.4	376	6	CA653223
C 460	19	1.5	1872	9	CG754248	533	18	1.4	376	6	CA653223
C 461	19	1.5	1872	9	CG754248	534	18	1.4	376	6	CA653223
C 462	19	1.5	1872	9	CG754248	535	18	1.4	376	6	CA653223

C 974	18	1.4	660	6	CA281099	SCFRSD102
C 975	18	1.4	660	8	A2974822	2M0249007
C 976	18	1.4	661	7	CN006615	CSBCS136D
C 977	18	1.4	662	5	BM114183	BM114183
C 978	18	1.4	662	2	BB640415	BB640415
C 979	18	1.4	663	6	CA296668	SCCFCR60
C 980	18	1.4	663	6	CB053220	NISC_g113
C 981	18	1.4	663	7	CK003436	AGENCOURT
C 982	18	1.4	663	9	CC895056	ZMBMB021
C 983	18	1.4	663	9	AG150068	Pan troy1
C 984	18	1.4	664	5	BM245160	BM245160
C 985	18	1.4	664	7	CK708311	ZF201-P00
C 986	18	1.4	664	8	AZ304487	1M0004003
C 987	18	1.4	664	8	BH871928	hn37E10.b
C 988	18	1.4	665	6	CA221326	SCSGFL403
C 989	18	1.4	665	6	CF201908	RR890915N
C 990	18	1.4	666	5	BM282854	BM282854
C 991	18	1.4	666	8	A0925780	RPCT-23-2
C 992	18	1.4	666	9	CC334921	CH240_413
C 993	18	1.4	667	5	BM274464	BM274464
C 994	18	1.4	667	8	BH001827	BMBAC01K1
C 995	18	1.4	668	4	BM157331	fv46h08.y
C 996	18	1.4	668	7	BM184567	fv69g07.y
C 997	18	1.4	668	7	CN632467	taf02a04.
C 998	18	1.4	668	9	CL811621	OR_CBA002
C 999	18	1.4	669	6	CA414369	UI-H-E20
C 1000	18	1.4	670	4	BJ803271	BJ803271

ALIGNMENTS

RESULT 1
BG470614
LOCUS
DEFINITION
60251153671 NIH_MGC_16 Homo sapiens CDNA clone IMAGE:4634169 5',
mRNA sequence.

ACCESSION
BG470614.1 GI:13402889

VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Cloned through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L10C1389 row: h column: 10
High quality sequence stop: 667.
Location/Qualifiers
1..730

FEATURES

source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4634169"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_16"
/note="Organ: eye; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACAGC(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit

ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 0.062; Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Note: this is a NIH_MGC Library."

OY
1110 TGGTGTGTTCCAGTGTGTTG 1135
|||||
665 TGGTGTGTTCCAGTGTGTTG 690

RESULT 2
AM193927/c
LOCUS
DEFINITION
AM193927 709 bp mRNA linear EST 29-NOV-1999
xmt19h03.x1 NCI CGAP U-4 Homo sapiens CDNA clone IMAGE:2684693 3',
similar to SW:TCBP2_HUMAN P40227 T-COMPLEX PROTEIN 1, ZETA SUBUNIT
;', mRNA sequence.

ACCESSION
AM193927
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
Cloned through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Seg primer: -40up from G1bco
High quality sequence stop: 405.
Location/Qualifiers
1..709

FEATURES

source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2684693"
/tissue_type="serous papillary carcinoma, high grade, 2 pooled tumors"
/lab_host="DH10B"
/clone_lib="NCI CGAP U-4"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.48 kb. Life Technologies catalog #: 11542-016"

ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 2.5; Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY
1113 TGGTGTGTTCCAGTGTGTTG 1135
|||||
698 TGGTGTGTTCCAGTGTGTTG 676

RESULT 3
LOCUS
DEFINITION
C1641530 1454 bp DNA linear GSS 06-JUL-2004
CH213-13H21.SP6 CH213 Gastrosteus aculeatus genomic clone
CH213-13H21 3', genomic survey sequence.

ACCESSION CL641530
 VERSION CL641530.1 GI:49660954
 KEYWORDS GSS.
 SOURCE Gasterosteus aculeatus (three spined stickleback)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes; Gasterosteidae; Gasterosteus.
 REFERENCE 1 (bases 1 to 1454)
 AUTHORS Kingsley, D., Grimwood, J., Dickson, M., Schmutz, J. and Myers, R.M.
 TITLE Expressed sequence tags from *Gasterosteus aculeatus*
 JOURNAL Unpublished (2004)
 COMMENT Contact: Grimwood, Jane
 Stanford Human Genome Center
 Stanford University School of Medicine
 975 S California Avenue, Palo Alto, CA 94304, USA
 Tel: 650 320 5917
 Fax: 650 320 5801
 Email: jane@hgsc.stanford.edu
 Plate: 13
 Class: BAC ends
 High quality sequence start: 30
 High quality sequence stop: 872.
 Location/Qualifiers
 1..1454
 /organism="Gasterosteus aculeatus"
 /mol_type="genomic DNA"
 /strain="Salmon River"
 /db_xref="taxon:69293"
 /clone="CH213-13H21"
 /sex="Mixed"
 /cell_type="Blood"
 /clone_1ib="CH213"
 /note="Vector: pTARBA2.1; Site 1: EcoRI. The sequence of the clone was established as a mapping and sequencing collaboration at the Stanford Genome Evolution Center, funded by the NIH Centers of Excellence in Genomic Science (CEGS) initiative (<http://cegs.stanford.edu>). The clone was isolated from the BAC library CHOR-213 built by Peter deJong in collaboration with the Stanford Genome Evolution Center (<http://www.chori.org/bacpac/>). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/ordering_information.htm)."

ORIGIN
 Query Match 1.8%; Score 23; DB 9; Length 1454;
 Best Local Similarity 100.0%; Pred. No. 2.6;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 89 CATTGTCATCAGATGCATGCA 111
 Db 961 CATTGTCATCAGATGCATGCA 983

RESULT 4
 LOCUS AQ663605 542 bp DNA 1linear GSS 23-JUN-1999
 DEFINITION HS_5476_A2_H04_T7A RPC1-11 Human Male BAC Library Homo sapiens genomic clone Plate=1052 Col=8 Row=O, genomic survey sequence.
 ACCESSION AQ663605
 VERSION AQ663605.1 GI:5171373
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 542)
 AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and Hood, L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 PUBMED 10449764
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong (pieter@dong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>
 Plate: 1052 Row: O Column: 8
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 542.
 Location/Qualifiers
 1..542
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="Plate=1052 Col=8 Row=O"
 /sex="male"
 /clone_1ib="RPC1-11 Human Male BAC Library"
 /note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

ORIGIN
 Query Match 1.7%; Score 22; DB 8; Length 542;
 Best Local Similarity 100.0%; Pred. No. 8.5;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 32 CATTCTCTGCGTGGATATTT 53
 Db 160 CATTCTCTGCGTGGATATTT 139

RESULT 5
 LOCUS BM705592 331 bp mRNA 1linear EST 28-FEB-2002
 DEFINITION UI-E-DWO-agi-c-04-0-UI.r1 UI-E-DWO Homo sapiens cDNA clone
 ACCESSION BM705592
 VERSION BM705592.1 GI:19018850
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 331)
 AUTHORS Bonaldi, M.F., Lennon, G. and Soares, M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@iowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.reegen.com).
 Seq primer: M13 Reverse.

FEATURES

source

Location/Qualifiers

1..331

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-E-DW0-agi-c-04-0-UI"

/tissue_type="lens"

/dev_stage="adult"

/lab_host="PH10B (Life Technologies) (T1 phage resistant)"

/clone_1ib="UI-E-DW0"

/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-E-DW0 is a cDNA library containing the following
 tissue(s): lens. The library was constructed according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was ligated to an EcoR I adaptor, digested with Not
 I, and cloned directionally into pT73-Pac vector. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)₁₈ tail. The
 sequence tag for this library is CATTAGCGA. This library
 was created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI)."

ORIGIN

Query Match 1.6%; Score 21; DB 4; Length 331;

Best Local Similarity 100.0%; Pred. No. 28;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTCTGCTG 1135

Db 55 GTGTGTTCCAGTCTGCTG 75

RESULT 6
 AV684301 349 bp mRNA linear EST 16-JAN-2002

LOCUS AV684301 GKC Homo sapiens cDNA clone GKCA0804 5', mRNA sequence.

ACCESSION AV684301

VERSION AV684301.1 GI:10286164

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, O., Cai, T., Zhang, X.,
 Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,
 Shen, K., Lu, G., Fu, G., Zhong, W., Xu, S., Gu, W., Huang, W., Zhao, X.,
 Hu, G., Gu, J., Chen, Z., and Han, Z.

TITLE Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

MEDLINE 21625106

COMMENT 11752456

CONTACT: Zeguang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919 (ex. 45)

Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn

This clone is available at CHGC in Shanghai.

Location/Qualifiers

1..349

/organism="Homo sapiens"

FEATURES

source

/mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GKCA0804"
 /tissue_type="hepatocellular carcinoma"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /clone_1ib="GKC"
 /note="Vector: pBluescript sk(-); Site 1: EcoRI; Site 2:
 XhoI"

ORIGIN

Query Match 1.6%; Score 21; DB 1; Length 349;

Best Local Similarity 100.0%; Pred. No. 28;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTCTGCTG 1135

Db 180 GTGTGTTCCAGTCTGCTG 200

RESULT 7
 AV696353 349 bp mRNA linear EST 16-JAN-2002

LOCUS AV696353 GKC Homo sapiens cDNA clone GKCAPB02 5', mRNA sequence.

ACCESSION AV696353

VERSION AV696353.1 GI:10298216

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, O., Cai, T., Zhang, X.,
 Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,
 Shen, K., Lu, G., Fu, G., Zhong, W., Xu, S., Gu, W., Huang, W., Zhao, X.,
 Hu, G., Gu, J., Chen, Z., and Han, Z.

TITLE Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

MEDLINE 21625106

COMMENT 11752456

CONTACT: Zeguang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919 (ex. 45)

Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn

This clone is available at CHGC in Shanghai.

Location/Qualifiers

1..349

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="GKCAPB02"

/tissue_type="hepatocellular carcinoma"

/dev_stage="Adult"

/lab_host="SOLR"

/clone_1ib="GKC"

/note="Vector: pBluescript sk(-); Site 1: EcoRI; Site 2:
 XhoI"

ORIGIN

Query Match 1.6%; Score 21; DB 1; Length 349;

Best Local Similarity 100.0%; Pred. No. 28;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTCTGCTG 1135

Db 180 GTGTGTTCCAGTCTGCTG 200

RESULT 8
AV697455 349 bp mRNA linear EST 16-JAN-2002
LOCUS AV697455 GKC Homo sapiens cDNA clone GKACPC02 5', mRNA sequence.
DEFINITION AV697455
ACCESSION AV697455
VERSION AV697455.1 GI:10299318
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 349)
AUTHORS Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,
Shen, K., Lu, G., Fu, G., Zhong, W., Xu, S., Gu, W., Huang, W., Zhao, X.,
Hu, G., Gu, J., Chen, Z. and Han, Z.
Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

JOURNAL MEDLINE 21625106
PUBMED 11752456
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers
1..349
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GKACPC02"
/issue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_1lb="GKC"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

ORIGIN
Query Match 1.6%; Score 21; DB 1; Length 349;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1115 GTGTGTTCCAGTCTGCTG 1135
|||||
160 GTGTGTTCCAGTCTGCTG 200

DB 160 GTGTGTTCCAGTCTGCTG 200

RESULT 9
AV695717 360 bp mRNA linear EST 16-JAN-2002
LOCUS AV695717 GKC Homo sapiens cDNA clone GKACZG01 5', mRNA sequence.
DEFINITION AV695717
ACCESSION AV695717
VERSION AV695717.1 GI:10297580
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 360)
AUTHORS Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,
Shen, K., Lu, G., Fu, G., Zhong, W., Xu, S., Gu, W., Huang, W., Zhao, X.,
Hu, G., Gu, J., Chen, Z. and Han, Z.
Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

MEDLINE 21625106
PUBMED 11752456
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers
1..360
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GKACZG01"
/issue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_1lb="GKC"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

ORIGIN
Query Match 1.6%; Score 21; DB 1; Length 360;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1115 GTGTGTTCCAGTCTGCTG 1135
|||||
160 GTGTGTTCCAGTCTGCTG 200

DB 160 GTGTGTTCCAGTCTGCTG 200

RESULT 10
B1036922/c 368 bp mRNA linear EST 14-JUN-2001
LOCUS B1036922
DEFINITION MR4-NT0141-080101-209-e04 NT0141 Homo sapiens cDNA, mRNA sequence.
ACCESSION B1036922
VERSION B1036922.1 GI:1443548
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 368)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagal, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V.,
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?c1=MR4&c2=MR4-NT0141-
080101-209-e04&c3=2001-01-08&c4=1)
Seq primer: puc 18 forward
High quality sequence start: 35
High quality sequence stop: 345.
Location/Qualifiers
1..368

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NT0141"
/note="Organ: nervous tumor; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN

Query Match 1.6%; Score 21; DB 4; Length 368;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1115 GTGTGTTCCAGTGCTGCTG 1135
|||||
135 GTGTGTTCCAGTGCTGCTG 115

RESULT 11 368 bp mRNA linear EST 04-MAR-2002
LOCUS BM762828
DEFINITION K-EST0044039 S12SN216 Homo sapiens cDNA clone S12SN216-12-C03 5',
mRNA sequence.
ACCESSION BM762828
VERSION BM762828.1 GI:19092443
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 368)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 12 row: C column: 03
High quality sequence stop: 368.
Location/Qualifiers

FEATURES
source 1..368
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S12SN216-12-C03"
/sex="F"
/tissue_type="lymph node"
/cell_type="Epithelial"
/cell_line="SNP-216"
/lab_host="Top10F"
/clone_lib="S12SN216"
/note="Organ: Stomach; Vector: pcns; Site 1: EcoRI;
Site 2: NotI; The poly (A) + RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand

converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

ORIGIN

Query Match 1.6%; Score 21; DB 4; Length 368;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1115 GTGTGTTCCAGTGCTGCTG 1135
|||||
209 GTGTGTTCCAGTGCTGCTG 229

RESULT 12 369 bp mRNA linear EST 13-MAR-1995
LOCUS F12095
DEFINITION HSC35F071 normalized infant brain cDNA Homo sapiens cDNA clone
c-35f07, mRNA sequence.
ACCESSION F12095
VERSION F12095.1 GI:706428
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 369)
AUTHORS Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
Devignes,M.D., Duprat,S., Houigatte,R., Jumeau,M.N., Lamy,B.,
Sebastiani-Kabatchis,C. and Tessier,A.

TITLE IMAGS: molecular integration of the analysis of the human genome
and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
PubMed 7757816
COMMENT Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.
Genexpress_library_id: C; Genexpress_sequence_id: y1c-35f07
Seq primer: (-21)M13 universal.
Location/Qualifiers

FEATURES
source 1..369
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="c-35f07"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/clone_lib="normalized infant brain cDNA"
/note="Organ: brain; Vector: lafmid BA; Site 1: HindIII;
Site 2: NotI; Sex=Female; dev stage=3 months old;
isolate=muscular atrophy patient; tissue type=total
brain; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Souares, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"

ORIGIN

Query Match 1.6%; Score 21; DB 7; Length 369;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1115 GTGTGTTCCAGTGCTGCTG 1135

Db 57 GTGGTTCAGTCTGCTGTG 77

RESULT 13
LOCUS BP623179/c 371 bp mRNA linear EST 26-JUN-2004
DEFINITION BP623179 RAF117 Arabidopsis thaliana cDNA clone RAF117-05-N11 3', mRNA sequence.

ACCESSION BP623179 GI:49274361
VERSION BP623179.1
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 371)

REFERENCE
AUTHORS Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T., Nakajima,M., Enju,A., Akiyama,K., Cono,Y., Muramatsu,M., Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K., Functional annotation of a full-length Arabidopsis cDNA collection Science 296 (5565), 141-145 (2002)
JOURNAL Science 296 (5565), 141-145 (2002)
MEDLINE 21932900
PUBMED 11910074
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: msek@rtc.riken.go.jp
reversed clone; please visit our web site
(http://p1sweb.gsc.riken.go.jp/) for further details.
Location/Qualifiers
1..371
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAF117-05-N11"
/lab_host="DH10B"
/clone_1fb="RAF117"
/note="Site 1: BamHI; Site 2: SalI; Subtraction library. The sequence was obtained from samples subjected to dehydration-treated (1, 2, 5, 10 and 24 hr) and rehydration-treated (1, 2, 5, 10, and 24 hr after dehydration treatment)."

ORIGIN

Query Match 1.6%; Score 21; DB 5; Length 371;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1126 GGTGCTGTGATACGATGCA 1146
|||||
280 GGTGCTGTGATACGATGCA 260

RESULT 14
LOCUS AW361182/c 372 bp mRNA linear EST 04-FEB-2000
DEFINITION RCI-CT0252-300999-011-c01 CT0252 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW361182
VERSION AW361182.1 GI:6865832
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 372)
HCGP http://www.judwig.org.br/ORESTES.
The FAPESP/LICR Human Cancer Genome Project

JOURNAL
COMMENT Unpublished (1999)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@judwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.judwig.org.br/scripts/gethtml2.pl?l=RC1&t2=RC1-CT0252-300999-011-c01&t3=1999-09-30&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 44
High quality sequence stop: 245.
Location/Qualifiers
1..372
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/def_stage="Adult"
/clone_1fb="CT0252"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 1.6%; Score 21; DB 2; Length 372;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGGTTCAGTCTGCTGTG 1135
|||||
132 GTGGTTCAGTCTGCTGTG 112

RESULT 15
LOCUS CF141588 386 bp mRNA linear EST 06-AUG-2003
DEFINITION UI-HP-CB0-aga-g-06-0-UI r1 NIH_MGC_210 Homo sapiens cDNA clone IMAGE:3099803 5', mRNA sequence.
ACCESSION CF141588
VERSION CF141588.1 GI:33257032
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 386)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Tim Ratliff
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/humanf1.html

FEATURES Seq primer: PYX-5.
Location/Qualifiers
source 1..386
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:309803"
/tissue_type="CNCAP(3)T-225 cell line"
/lab_host="DH10B (T1 phage resistant)"
/clone_11b="NIH MGc 210"
/note="Organ: Prostate; Vector: pT73 Pac; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dt primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction. Ligated with EcoR I adaptor, digested with Not I and then cloned directionally into pT73 Pac vector. The library tag sequence located between the Not I site and the polyA tail is CCCAC. Tissue was provided by Tim Ratliff."

Query Match 1.6%; Score 21; DB 6; Length 386;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1115 GTGTGTTCCAGTGCTGTG 1135
|||||
Db 262 GTGTGTTCCAGTGCTGTG 282
|||||

RESULT 16
LOCUS CN345987 414 bp mRNA linear EST 16-MAY-2004
DEFINITION 1700600003542 GRN_PRENEM Homo sapiens cDNA 5', mRNA sequence.
ACCESSION CN345987
VERSION CN345987.1 GI:47345921
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 414)
Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J., Li, Y., Xu, C., Pang, R., Guegler, K., Rao, M.S., Mandalam, R., Lebkowski, J. and Stanton, L.W.
Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation
Nat. Biotechnol. 22 (6), 707-716 (2004)
Contact: Brandenberger R
Regenerative Medicine
Geron Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbrandenberger@geron.com
Insert Length: 414 Std Error: 0.00.
Location/Qualifiers
1..414
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cell, retinoic acid and mitogen-treated hES cell line H7"
/clone_11b="GRN_PRENEM"
/note="Oligo dT primed, full-length enriched cDNA library from hES cell line H7 (p29) maintained in feeder-free conditions. Embryoid bodies were generated in the presence of all-trans retinoic acid and mitogens."

ORIGIN
Query Match 1.6%; Score 21; DB 7; Length 414;

Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1115 GTGTGTTCCAGTGCTGTG 1135
|||||
Db 359 GTGTGTTCCAGTGCTGTG 379
|||||

RESULT 17
LOCUS AA150852 424 bp mRNA linear EST 19-MAY-1997
DEFINITION 144405.x1 Soares pregnant uterus NBHPU Homo sapiens cDNA clone IMAGE:504753 5', similar to TR:G184462 G184462 CHAPERONIN-LIKE PROTEIN, ;, mRNA sequence.
ACCESSION AA150852
VERSION AA150852.1 GI:1722363
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 424)
Hillier, L., Lennon, G., Becker, M., Bonaldi, M.F., Chiappelli, B., Chisoe, S., Dietrich, N., Dubuque, T., Favvello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
MEDLINE 97044478
PUBMED 8889549
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 807 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham.
Location/Qualifiers
1..424
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3809741"
/db_xref="taxon:9606"
/clone="IMAGE:504753"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_11b="Soares pregnant uterus NBHPU"
/note="Organ: uterus; Vector: pT73-Pac; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGAGAGATTCGCGCGCCCTTTTCTTTTCTTTTCTTTT 3', double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldi."

ORIGIN
Query Match 1.6%; Score 21; DB 1; Length 424;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1115 GTGTGTTCCAGTGCTGTG 1135
|||||
Db 82 GTGTGTTCCAGTGCTGTG 102
|||||

RESULT 18
 W69312 425 bp mRNA linear EST 16-OCT-1996
 LOCUS z44601.r1 Soares fetal heart NbH19W Homo sapiens cDNA clone
 DEFINITION IMAGE:343728.5, similar to PIR:A46141 A46141 chaperonin-like protein HTR3 - human ;, mRNA sequence.
 W69312
 ACCESSION W69312.1 GI:1378573
 VERSION EST.
 KEYWORDS Homo sapiens (human).
 SOURCE Homo sapiens (human).
 ORGANISM Homo sapiens (human).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 425)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, P., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1259 Std Error: 0.00
 Seq primer: mob.REGA+ET
 High quality sequence stop: 240.
 Location/Qualifiers
 1..425
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="Gene:1269103"
 /db_xref="taxon:9606"
 /clone="IMAGE:343728"
 /sex="unknown"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_1ib="Soares fetal heart NbH19W"
 /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5'-TGTACCAATCTGAGAGTGGAGCGCCGATCTTTTCTTTTCTTTT 3'), double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same focus as the fetal lung library, Soares fetal lung NbH19W."

KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 426)
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.U., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 PUBMED 10737800
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the PAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM3&ct=CM3-HT0193-031199-030-f02&ts=1999-11-03&ct=4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 36
 High quality sequence stop: 426.
 Location/Qualifiers
 1..426
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_1ib="HT0193"
 /note="Organ: head neck; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN
 Query Match 1.6%; Score 21; DB 5; Length 426;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTCTGTG 1135
 |||||
 Db 160 GTGTGTTCCAGTCTGTG 140

RESULT 20
 T66047 426 bp mRNA linear EST 07-MAR-1995
 LOCUS YC79e07.r1 Soares infant brain JN1B Homo sapiens cDNA clone
 DEFINITION IMAGE:21912.5, similar to SP:A46141 A46141 CHAPERONIN-LIKE-HTR3 - ;, mRNA sequence.
 T66047
 ACCESSION T66047.1 GI:675092
 VERSION EST.
 KEYWORDS Homo sapiens (human).
 SOURCE Homo sapiens (human).
 ORGANISM Homo sapiens (human).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 426)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

TITLE Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 JOURNAL Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
 COMMENT Wilson, R.
 The WashU-Merck EST Project
 Unpublished (1995)
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

Insert Size: 1405
 High quality sequence stops: 362 Source: IMAGE Consortium, LNL This
 clone is available royalty-free through LNL; contact the IMAGE
 Consortium (info@image.llnl.gov) for further information.
 Insert length: 1405 Std Error: 0.00
 Seq primer: M13Rpi
 High quality sequence stop: 362.
 Location/Qualifiers

FEATURES

source

```
1. 426
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:394259"
/db_xref="taxon:9606"
/clone="IMAGE:21912"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares infant brain LNB"
/note="Organ: whole brain. Vector: lafmid BA; Site 1: Not
1; Site 2: Hind III; 1st strand cDNA was primed with a Not
1 - oligo(dT) primer [5']
AAGCGAAGATTGCGCGCCGACGAATTTTTTTTTTTT 3';
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the lafmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."
```

ORIGIN

Query Match 1.6%; Score 21; DB 7; Length 426;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1115 GTGTGTTCCAGTGTCTGTG 1135
 ||||||||||||||||||||
 DB 28 GTGTGTTCCAGTGTCTGTG 48

RESULT 21
 BM853177 434 bp mRNA linear EST 06-MAR-2002
 LOCUS K-EST0134548 S22SNU16n1 Homo sapiens cDNA clone S22SNU16n1-134-C12
 DEFINITION 5', mRNA sequence.
 ACCESSION BM853177
 VERSION BM853177.1 GI:19209576
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 434)
 Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
 Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
 Kim, Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished (2002)
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409

FEATURES
 source
 1. 434
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S22SNU16n1-134-C12"
 /sex="F"
 /tissue_type="Asciates"
 /cell_type="Lymphoblast-like"
 /cell_line="SNU-16"
 /lab_host="DH10B"
 /clone_lib="S22SNU16n1"
 /note="Organ: Stomach; Vector: pT73-Pac; Site 1: EcoRI;
 Site 2: NotI; The S22SNU16 library was contributed by
 Soares Laboratory and it was constructed as described by
 Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome
 Research 6(9): 791-806. RNA was prepared from harvested
 cells of SNU-16 culture. SNU-16 cell was obtained from
 Korean Cell Line Bank (KCLB). SNU-16 was established from
 ascitic fluids of Korean patients by Park J.G. et al.
 (1990), Cancer Res 50: 2773-2780."

ORIGIN

Query Match 1.6%; Score 21; DB 4; Length 434;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1115 GTGTGTTCCAGTGTCTGTG 1135
 ||||||||||||||||||||
 DB 371 GTGTGTTCCAGTGTCTGTG 391

RESULT 22
 CB117296 434 bp mRNA linear EST 28-JAN-2003
 LOCUS K-EST0162463 L4SNU368 Homo sapiens cDNA clone L4SNU368-26-H04 5',
 DEFINITION mRNA sequence.
 ACCESSION CB117296
 VERSION CB117296.1 GI:27943103
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 434)
 Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
 Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
 Kim, Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished (2002)
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 26 row: H column: 04
 High quality sequence stop: 434.
 Location/Qualifiers

FEATURES

source

```
1. 434
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L4SNU368-26-H04"
/sex="M"
/tissue_type="Liver"
/cell_type="Polygonal"
/cell_line="SNU-368"
/lab_host="Top10F"
```


/clone_11b="L4SNU368"
 /note="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI;
 Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then deacapped
 with tobacco acid pyrophosphatase (TAP). The deacapped
 intact mRNA was ligated with DNA-RNA linker including
 EcoRI site by treatment of T4 RNA ligase and the first
 strand cDNA was synthesized from oligo dT-selected mRNA by
 priming with dT-tailed vector. The dT-tailed vector was
 adjusted to have about 60nt. The cDNA vector was
 circularized with E. coli DNA ligase after digestion of
 EcoRI which site is also included in vector. An RNA strand
 converted to a DNA strand by Okayama-Berg method. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F' by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."

ORIGIN

Query Match 1.6%; Score 21; DB 6; Length 434;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1115 GTGTGTTCCAGTGTGCTG 1135
 ||||||||||||||||
 Db 155 GTGTGTTCCAGTGTGCTG 175

RESULT 23
 A1910561 441 bp mRNA linear EST 28-JUL-1999
 LOCUS Wf88h07.x1 Soares NFL T GBC S1 Homo sapiens cDNA clone
 DEFINITION IMAGE:2362525 3' similar to contains Alu repetitive
 element;contains element MER28 repetitive element;; mRNA sequence.
 A1910561
 A1910561.1 GI:5630297

ACCESSION
 VERSION A1910561.1 GI:5630297
 SOURCE EST.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 441)
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@db-rcmail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40UP from Gibco
 High quality sequence stop: 440.
 Location/Qualifiers
 1..441

FEATURES
 source
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2362525"
 /lab_host="DH10B"
 /note="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI;
 Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then deacapped
 with tobacco acid pyrophosphatase (TAP). The deacapped
 intact mRNA was ligated with DNA-RNA linker including
 EcoRI site by treatment of T4 RNA ligase and the first
 strand cDNA was synthesized from oligo dT-selected mRNA by
 priming with dT-tailed vector. The dT-tailed vector was
 adjusted to have about 60nt. The cDNA vector was
 circularized with E. coli DNA ligase after digestion of
 EcoRI which site is also included in vector. An RNA strand
 converted to a DNA strand by Okayama-Berg method. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F' by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."

FEATURES

source

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2362525"
 /lab_host="DH10B"
 /note="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI;
 Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then deacapped
 with tobacco acid pyrophosphatase (TAP). The deacapped
 intact mRNA was ligated with DNA-RNA linker including
 EcoRI site by treatment of T4 RNA ligase and the first
 strand cDNA was synthesized from oligo dT-selected mRNA by
 priming with dT-tailed vector. The dT-tailed vector was
 adjusted to have about 60nt. The cDNA vector was
 circularized with E. coli DNA ligase after digestion of
 EcoRI which site is also included in vector. An RNA strand
 converted to a DNA strand by Okayama-Berg method. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F' by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."

ORIGIN

Query Match 1.6%; Score 21; DB 1; Length 441;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1260 TGTACTTGAGAAAAACCGA 1280
 ||||||||||||||||
 Db 335 TGTACTTGAGAAAAACCGA 355

RESULT 24
 A1127935 443 bp mRNA linear EST 26-NOV-1996
 LOCUS 214f07.x1 Soares pregnant uterus NBHPU Homo sapiens cDNA clone
 DEFINITION IMAGE:501925 5' similar to TR:0187462 G184462 CHAPERONIN-LIKE
 PROTEIN;; mRNA sequence.
 A1127935
 A1127935.1 GI:1687363

ACCESSION
 VERSION A1127935.1 GI:1687363
 SOURCE EST.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 443)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawking, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.
 The WashU-Merck EST Project
 Unpublished (1995)
 Contact: Wilson RK
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28M3 rev2 from Amerham
 High quality sequence stop: 311.
 Location/Qualifiers
 1..443

FEATURES

source

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="GDB:3806913"
 /db_xref="taxon:9606"
 /clone="IMAGE:501925"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_11b="Soares pregnant uterus NBHPU"
 /note="Organ: uterus; Vector: pT73-Pac; Site 1: Not I;
 Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
 oligo (dT) primer (5',
 AACTGAGAAATTCGCGCGCTTTTCTTTTCTTTTCTTTT 3'),
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by M. Fatima Bonaldo."

ORIGIN

Query Match 1.6%; Score 21; DB 1; Length 443;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1115 GTGTGTTCCAGTGTGCTG 1135
 ||||||||||||||||
 Db 82 GTGTGTTCCAGTGTGCTG 102

RESULT 25

BG761673 459 bp mRNA linear EST 15-MAY-2001
 LOCUS 602718849P1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4858669 5',
 DEFINITION mRNA sequence.
 BG761673
 ACCESSION BG761673.1 GI:14072326
 VERSION
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 459)
 NIH-MGC http://img.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgabs-rc@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/PTP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1CM1713 row: b column: 14
 High quality sequence stop: 457.
 Location/Qualifiers
 1..459
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4858669"
 /issue_type="melanotic melanoma, high MDR (cell line)"
 /lab_host="MD10B (phage-resistant)"
 /clone_lib="NIH_MGC_49"
 /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; CDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGACGAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-CDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 library."

ORIGIN
 Query Match 1.6%; Score 21; DB 4; Length 459;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGGTCTGTG 1135
 |||||||
 Db 107 GTGTGTTCCAGGTCTGTG 127

RESULT 26 493 bp mRNA linear EST 21-JUL-2000
 BE392663
 LOCUS 601307517F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3625229 5',
 DEFINITION mRNA sequence.
 BE392663
 ACCESSION BE392663.1 GI:9338028
 VERSION
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 493)
 NIH-MGC http://img.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgabs-rc@mail.nih.gov

Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1CM305 row: a column: 06
 High quality sequence stop: 493.
 Location/Qualifiers
 1..493
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3625229"
 /issue_type="endometrium, adenocarcinoma cell line"
 /lab_host="MD10B (phage-resistant)"
 /clone_lib="NIH_MGC_44"
 /note="Organ: uterus; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; CDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-CDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN
 Query Match 1.6%; Score 21; DB 2; Length 493;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGGTCTGTG 1135
 |||||||
 Db 330 GTGTGTTCCAGGTCTGTG 350

RESULT 27 499 bp mRNA linear EST 06-MAR-2002
 BM641588
 LOCUS K-EST0118850 S12SNW216 Homo sapiens cDNA clone S12SNW216-53-B11 5',
 DEFINITION mRNA sequence.
 BM641588
 ACCESSION BM641588.1 GI:19197997
 VERSION
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 499)
 KIM,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.W., Park,H.S., Kim,S. and
 Kim,Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished (2002)
 CONTACT: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongseung@mail.kribb.re.kr
 Plate: 53 row: B column: 11
 High quality sequence stop: 499.
 Location/Qualifiers
 1..499
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S12SNW216-53-B11"
 /sex="F"
 /issue_type="lymph node"
 /cell_type="Epithelial"
 /cell_line="SNU-216"

/lab_host="Top10P"
/clone_1lb="S1S2NU216"
/note="Organ: Stomach; Vector: pcn5; Site 1: EcoRI;
Site 2: NotI; The poly (A) + RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10P by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

ORIGIN

Query Match 1.6%; Score 21; DB 4; Length 499;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1115 GTGTGTTCCAGTGTCTGTG 1135
|||||
Db 209 GTGTGTTCCAGTGTCTGTG 229

RESULT 28
BE265418 504 bp mRNA linear EST 13-JUL-2000
LOCUS 601194958P1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:353856 5',
DEFINITION mRNA sequence.
ACCESSION BE265418
VERSION BE265418.1 GI:9138986
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 504)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Plate: L1CM224 row: d column: 15
High quality sequence stop: 504.
Location/Qualifiers

FEATURES

source
1..504
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3538526"
/issue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_7"
/note="Organ: lung; Vector: pOTR1; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAACAG(C). Size-selected >800bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Query Match 1.6%; Score 21; DB 2; Length 504;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1115 GTGTGTTCCAGTGTCTGTG 1135
|||||
Db 67 GTGTGTTCCAGTGTCTGTG 87

RESULT 29
CN345984 506 bp mRNA linear EST 16-MAY-2004
LOCUS 17000600169383 GRN_PRENBU Homo sapiens cDNA 5', mRNA sequence.
DEFINITION CN345984
ACCESSION CN345984
VERSION CN345984.1 GI:47345918
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 506)
Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J.,
Li, Y., Xu, C., Pang, R., Guejler, K., Rao, M.S., Mandalam, R.,
Lebkowski, J. and Stanton, L.W.
Transcriptome characterization elucidates signaling networks that
control human ES cell growth and differentiation
Nat. Biotechnol. 22 (6), 707-716 (2004)
Contact: Brandenberger R
Regenerative Medicine
Geron Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbrandenberger@geron.com
Insert Length: 506 Std Error: 0.00.
Location/Qualifiers

FEATURES

source
1..506
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/issue_type="embryonic stem cell, retinoic acid and
mitogen-treated hES cell line H7"
/clone_1lb="GRN_PRENBU"
/note="Oligo dt-primed, full-length enriched cDNA library
from hES cell line H7 (p29) maintained in feeder-free
conditions. Embryoid bodies were generated in the presence
of all-trans retinoic acid and mitogens."

ORIGIN

Query Match 1.6%; Score 21; DB 7; Length 506;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1115 GTGTGTTCCAGTGTCTGTG 1135
|||||
Db 466 GTGTGTTCCAGTGTCTGTG 506

RESULT 30
CB159634 511 bp mRNA linear EST 30-JAN-2003
LOCUS K-EST0219212 L18POOL1n1 Homo sapiens cDNA clone L18POOL1n1-25-B04
DEFINITION 5', mRNA sequence.
ACCESSION CB159634
VERSION CB159634.1 GI:28145760
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 511)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.
Oh, K.J., Cheong, J.E., Sohn, H.Y., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Kim, Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS

Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 25, row: B column: 04
High quality sequence stop: 511.
Location/Qualifiers
1. 511

FEATURES
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="U18POOL1n1-25-B04"
/cell_line="SNU-354+Cho-CK+Choi-CK+HLK-3"
/lab_host="Top10F"
/clone_lib="U18POOL1n1"
/note="Organ: Liver; Vector: pRTT3-Pac; Site_1: EcoRI;
Site_2: NotI; The library was contributed by the Soares
Laboratory and it was constructed as described by Bonaldo,
M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
6(9): 791-806. RNA was prepared from harvested cell
culture."

ORIGIN
Query Match 1.6%; Score 21; DB 6; Length 511;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTGTCTGTG 1135
|||||
Db 443 GTGTGTTCCAGTGTCTGTG 463

RESULT 31
BM796788 516 bp mRNA linear EST 05-MAR-2002
LOCUS K-EST0079591 S22SNU16n1 Homo sapiens cDNA clone S22SNU16n1-63-H11
DEFINITION 5', mRNA sequence.
ACCESSION BM796788
VERSION BM796788.1 GI:19145020
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 516)
Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.U., Cheong, J.E., Sohn, H.Y., Kim, J.W., Park, H.S., Kim, S. and
Kim, Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS

TITLE
JOURNAL
COMMENT
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 63 row: H column: 11
High quality sequence stop: 516.
Location/Qualifiers
1. 516

FEATURES
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S22SNU16n1-63-H11"
/sex="F"
/tissue_type="Ascites"
/cell_type="Lymphoblast-like"
/cell_line="SNU-16"
/lab_host="DH10B"
/clone_lib="S22SNU16n1"

/note="Organ: Stomach; Vector: pRTT3-Pac; Site_1: EcoRI;
Site_2: NotI; The S22SNU16 library was contributed by the
Soares laboratory and it was constructed as described by
Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome
Research 6(9): 791-806. RNA was prepared from harvested
cells of SNU-16 culture. SNU-16 cell was obtained from
Korean Cell Line Bank (KCLB). SNU-16 was established from
ascitic fluids of Korean patients by Park J.G. et al.
(1990), Cancer Res 50: 2773-2780."

ORIGIN
Query Match 1.6%; Score 21; DB 4; Length 516;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTGTCTGTG 1135
|||||
Db 321 GTGTGTTCCAGTGTCTGTG 341

RESULT 32
BG149625 522 bp mRNA linear EST 05-FEB-2001
LOCUS nad31d12.x1 NCI CGAP Lu24 Homo sapiens cDNA IMAGE:3366983 3'
DEFINITION similar to SW:TCPEZ_HUMAN P40227 T-COMPLEX PROTEIN 1, ZETA SUBUNIT
; , mRNA sequence.
ACCESSION BG149625
VERSION BG149625.1 GI:12661655
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 522)
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapb-rt@mail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL, send email to:
info@image.lnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 460.
Location/Qualifiers
1. 522

FEATURES
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3366983"
/tissue_type="carcinoid"
/lab_host="DH10B"
/clone_lib="NCI-CGAP Lu24"
/note="Organ: lung; Vector: pRTT3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI CGAP Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (clonids
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 1.6%; Score 21; DB 4; Length 522;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTGTCTGTG 1135
 |||||
 DB 458 GTGTGTTCCAGTGTCTGTG 478

RESULT 33
 CB124236 528 bp mRNA linear EST 29-JAN-2003
 LOCUS K-EST0172905 L5HLK181 Homo sapiens cDNA clone L5HLK181-21-B04 5',
 DEFINITION mRNA sequence.

ACCESSION CB124236
 CB124236
 VERSION CB124236.1 GI:28084249
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 528)
 Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.B., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001
 JOURNAL Unpublished (2002)
 COMMENT Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 21 row: B column: 04
 High quality sequence stop: 528.

FEATURES

source
 1..528
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="L5HLK181-21-B04"
 /sex="M"
 /cell_line="HLK-1"
 /lab_host="Top10P"
 /clone_1ib="L5HLK181"
 /note="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI;
 Site 2: NotI; The poly (A) + RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then deacapped
 with tobacco acid pyrophosphatase (TAP). The deacapped
 intact mRNA was ligated with DNA-RNA linker including
 EcoRI site by treatment of T4 RNA ligase and the first
 strand cDNA was synthesized from oligo dT-selected mRNA by
 priming with dt-tailed vector. The dt-tailed vector was
 adjusted to have about 60nt. The cDNA vector was
 circularized with E. coli DNA ligase after digestion of
 EcoRI which site is also included in vector. An RNA strand
 converted to a DNA strand by Okayama-Berg method. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10P' by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library. After analyzing and
 sequencing about 2,000 - 3,000 colonies in original cDNA
 library, the abundant cDNAs were selected and amplified by
 PCR reaction using vector region primer including T7
 promoter as 5' primer and N(dT)14 as 3' primer. The PCR
 products were used as template for synthesis of
 biotinylated single stranded RNA by in vitro transcription
 reaction. The synthesized RNA probes were hybridized with
 antisense single stranded cDNAs prepared from original
 library and incubated with avidin-gel. After removing
 DNA-RNA hybrids by centrifuge, the subtracted cDNA
 libraries were constructed by transformation of the
 remaining DNA into competent cells E. coli Top10P' with
 electroporation method."

ORIGIN

Query Match 1.6%; Score 21; DB 6; Length 528;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTGTCTGTG 1135
 |||||
 DB 343 GTGTGTTCCAGTGTCTGTG 363

RESULT 34
 W38880
 LOCUS

DEFINITION zb32a05.r1 Soares parathyroid tumor NbHPA Homo sapiens cDNA clone
 IMAGE:305264 5' similar to PIR:A46141 A46141 chaperonin-1like
 protein HTR3 - human ;, mRNA sequence.

ACCESSION W38880
 W38880
 VERSION W38880.1 GI:1320586
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 536)
 Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
 Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
 Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
 Trevaaskis,E., Waterston,R., Willimsom,A., Wohlmann,P. and
 Wilson,R.

TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watsn.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: mob.REGA+ET
 High quality sequence stop: 383.

FEATURES

source
 1..536
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="GDB:1248674"
 /db_xref="taxon:9606"
 /clone="IMAGE:305264"
 /tissue="parathyroid tumor"
 /dev_stage="adult"
 /lab_host="PH10B (ampicillin resistant)"
 /clone_1ib="Soares_parathyroid_tumor_NbHPA"
 /note="Organ: parathyroid gland; Vector: pT73D
 (pharmacia) with a modified polylinker; Site 1: Not I;
 Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
 oligo(dT) primer
 [5'-GTGTACAACTCTGAAGTGGAGCGCGCCACACATTTTCTTTTCTTTT
 TTTT-3'] , double-stranded cDNA was size selected, ligated
 to Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of a modified pT73
 vector (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M.Patino Bonaldi. RNA from sporadic parathyroid
 adenomas was kindly provided by Dr. Stephen Marx, National
 Institute of Diabetes and Digestive and Kidney Diseases,
 NIH."

ORIGIN

Query Match 1.6%; Score 21; DB 7; Length 536;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTGTCTGTG 1135

ORIGIN

Db 79 GTGTGTTCCAGGTGCTGATG 99

RESULT 35
BM840031

LOCUS
DEFINITION BM840031 542 bp mRNA linear EST 06-MAR-2002
K-EST0117031 S12SNU216 Homo sapiens cDNA clone S12SNU216-31-B03 5',
mRNA sequence.

ACCESSION
BM840031

VERSION
BM840031.1 GI:19196440

KEYWORDS
EST.

SOURCE
Homo sapiens (human)

ORGANISM
Homo sapiens

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 542)

AUTHORS
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.

TITLE
21C Frontier Korean EST Project 2001

JOURNAL
Unpublished (2002)

COMMENT
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 31 row: B column: 03
High quality sequence stop: 542.

FEATURES
source
1..542
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S12SNU216-31-B03"
/sex="F"
/tissue_type="lymph node"
/cell_line="Sp1hel1a1"
/cell_line="SNU-216"
/lab_host="Top10F"
/lab_host="S12SNU216"
/note="Organ: Stomach; Vector: pCNS; Site 1: EcoRI;
Site 2: NotI; The poly (A) + RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then deacapped
with tabacco acid pyrophosphatase (TAP). The deacapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by electroporation method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

ORIGIN

Query Match 1.6%; Score 21; DB 4; Length 542;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGGTGCTGATG 1135
|||||
209 GTGTGTTCCAGGTGCTGATG 229

RESULT 36
CB119331

LOCUS
CB119331 543 bp mRNA linear EST 28-JAN-2003

DEFINITION K-EST0166075 L10choCK0 Homo sapiens cDNA clone L10choCK0-9-B04 5',
mRNA sequence.

ACCESSION
CB119331

VERSION
CB119331.1 GI:27945135

KEYWORDS
EST.

SOURCE
Homo sapiens (human)

ORGANISM
Homo sapiens

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 543)

AUTHORS
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.

TITLE
21C Frontier Korean EST Project 2001

JOURNAL
Unpublished (2002)

COMMENT
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 9 row: B column: 04
High quality sequence stop: 543.

FEATURES
source
1..543
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L10choCK0-9-B04"
/sex="M"
/cell_line="Cho-CK"
/lab_host="Top10F"
/lab_host="L10choCK0"
/note="Organ: Liver; Vector: pT73-Pac; Site 1: EcoRI;
Site 2: NotI; The library was contributed by the Soares
Laboratory and it was constructed as described by Bonaldo,
M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
6(9): 791-806. RNA was prepared from harvested cell
culture."

ORIGIN

Query Match 1.6%; Score 21; DB 6; Length 543;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGGTGCTGATG 1135
|||||
505 GTGTGTTCCAGGTGCTGATG 525

RESULT 37
CB147674

LOCUS
CB147674 547 bp mRNA linear EST 29-JAN-2003

DEFINITION K-EST0203727 L14choiCK0 Homo sapiens cDNA clone L14choiCK0-17-G06
5', mRNA sequence.

ACCESSION
CB147674

VERSION
CB147674.1 GI:28129145

KEYWORDS
EST.

SOURCE
Homo sapiens (human)

ORGANISM
Homo sapiens

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 547)

AUTHORS
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.

TITLE
21C Frontier Korean EST Project 2001

JOURNAL
Unpublished (2002)

COMMENT
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 17 row: G column: 06
 High quality sequence stop: 547.
 Location/Qualifiers

/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Umbilical vein endothelial cells, passaged once. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAAATCGGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'"

FEATURES
 source
 1..547
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="U14choiCK0-17-G06"
 /sex="M"
 /cell_line="Choi-CK"
 /lab_host="Top10F"
 /clone_lib="U14choiCK0"
 /note="Organ: Liver; Vector: pT7T3-Pac; Site 1: EcoRI; Site 2: NotI; The library was contributed by the Soares laboratory and it was constructed as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from harvested cell culture."

ORIGIN

Query Match 1.6%; Score 21; DB 6; Length 547;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTCTGTGTG 1135
 |||||
 Db 388 GTGTGTTCCAGTCTGTGTG 408

RESULT 38
 AA186456 548 bp mRNA linear EST 10-MAR-1998
 LOCUS zpt0hio.r1 Stratsgene endothelial cell 937223 Homo sapiens cDNA
 DEFINITION clone IMAGE:625603 5' similar to TR:G184462 G184462 CHAPERONIN-LIKE
 PROTEIN. // mRNA sequence.
 ACCESSION AA186456
 VERSION AA186456.1 GI:1774556
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 548)
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,
 Kitzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
 Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
 Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
 JOURNAL WASHU-NCI human EST Project
 COMMENT Unpublished (1997)
 CONTACT: Wilson RK
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: eschwartzon.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGS Consortium (info@image.lnl.gov) for further information.
 Insert Length: 931 Std Error: 0.00
 Seq primer: -28M13 rev2 from Amerham
 High quality sequence stop: 301.
 Location/Qualifiers
 1..548
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="GDB:5047757"
 /db_xref="taxon:9606"
 /clone="IMAGE:625603"
 /dev_stage="umbilical vein, 1 passage"
 /lab_host="SOLR (kanamycin resistant)"
 /clone_lib="Stratsgene endothelial cell 937223"

ORIGIN
 Query Match 1.6%; Score 21; DB 1; Length 548;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTCTGTGTG 1135
 |||||
 Db 112 GTGTGTTCCAGTCTGTGTG 132

RESULT 39
 CB160975 553 bp mRNA linear EST 30-JAN-2003
 LOCUS K-EST0220750 L18POOL1n1 Homo sapiens cDNA clone L18POOL1n1-24-B06
 DEFINITION 5' mRNA sequence.
 ACCESSION CB160975
 VERSION CB160975.1 GI:28147101
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 553)
 AUTHORS Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
 Oh, K.U., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
 Kim, Y.S.
 JOURNAL 21C Frontier Korean EST Project 2001
 COMMENT Unpublished (2002)
 CONTACT: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 24 row: B column: 06
 High quality sequence stop: 553.
 Location/Qualifiers
 1..553
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="L18POOL1n1-24-B06"
 /cell_line="SNU-354+choi-CK+HLK-3"
 /lab_host="Top10F"
 /clone_lib="L18POOL1n1"
 /note="Organ: Liver; Vector: pT7T3-Pac; Site 1: EcoRI; Site 2: NotI; The library was contributed by the Soares laboratory and it was constructed as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from harvested cell culture."

ORIGIN
 Query Match 1.6%; Score 21; DB 6; Length 553;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GTGTGTTCCAGTCTGTGTG 1135
 |||||
 Db 427 GTGTGTTCCAGTCTGTGTG 447

RESULT 40
 AO837961/c 553 bp DNA linear GSS 30-AUG-1999
 LOCUS AO837961


```

DEFINITION HS_4647_B2_A06_SPEE CIT Approved Human Genomic Sperm Library D Homo
             sapiens genomic clone Plate=4647 Col=12 Row=B, genomic survey
sequence.
ACCESSION   AO837961
VERSION     AO837961
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (bases 1 to 553)
            Mahairas,G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
            Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
            Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
JOURNAL     MEDLINE
PUBMED      10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 4647 row: B column: 12
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 553.
FEATURES
    source
        1..553
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /clone="plate=4647 Col=12 Row=B"
            /sex="male"
            /clone_lib="CIT Approved Human Genomic Sperm Library D"
            /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
            E-Coli DH10B"
ORIGIN
Query Match      1.6%; Score 21; DB 8; Length 553;
Best Local Similarity 100.0%; Pred.No. 23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      192 ATATCATATCATCAATAC 212
        ||||||||||||||||
Db      543 ATATCATATCATCAATAC 523

Search completed: January 26, 2005, 10:01:08
Job time : 3136 secs

```



```
977 15 1.2 3585 1 US-08-455-543A-34 Sequence 34, Appl
978 15 1.2 3585 2 US-08-193-078B-23 Sequence 23, Appl
979 15 1.2 3585 2 US-08-223-305C-34 Sequence 34, Appl
980 15 1.2 3585 3 US-08-949-386-30 Sequence 30, Appl
981 15 1.2 3585 3 US-08-450-562-30 Sequence 30, Appl
982 15 1.2 3585 3 US-08-984-709A-30 Sequence 30, Appl
983 15 1.2 3585 3 US-08-450-272-30 Sequence 30, Appl
984 15 1.2 3585 4 US-08-450-272-30 Sequence 30, Appl
985 15 1.2 3585 4 US-08-450-272-30 Sequence 4259, Ap
986 15 1.2 3600 1 US-08-455-543A-11 Sequence 11, Appl
987 15 1.2 3600 1 US-08-193-078B-11 Sequence 33, Appl
988 15 1.2 3600 2 US-08-223-305C-11 Sequence 11, Appl
989 15 1.2 3600 2 US-08-223-305C-11 Sequence 33, Appl
990 15 1.2 3600 2 US-08-149-097D-11 Sequence 11, Appl
991 15 1.2 3600 3 US-08-949-386-11 Sequence 11, Appl
992 15 1.2 3600 3 US-08-450-562-11 Sequence 11, Appl
993 15 1.2 3600 3 US-08-984-709A-11 Sequence 11, Appl
994 15 1.2 3600 3 US-08-450-272-11 Sequence 11, Appl
995 15 1.2 3600 3 US-08-984-709A-11 Sequence 11, Appl
996 15 1.2 3600 4 US-08-450-272-11 Sequence 11, Appl
997 15 1.2 3601 3 US-09-017-631-23 Sequence 23, Appl
998 15 1.2 3602 2 US-08-883-795A-33 Sequence 33, Appl
999 15 1.2 3602 3 US-09-018-138-1 Sequence 1, Appl
1000 15 1.2 3628 3 US-08-480-640A-113 Sequence 113, Appl
```

ALIGNMENTS

```
RESULT 1
US-09-792-024-42
/ Sequence 42, Application US/09792024
/ Patent No. 6783985
/ GENERAL INFORMATION:
/ APPLICANT: Roemer, Terry
/ APPLICANT: Jiang, Bo
/ APPLICANT: Boone, Charles
/ APPLICANT: Bussey, Howard
/ TITLE OF INVENTION: Gene Disruption Methodologies for Drug
/ FILE REFERENCE: 10187-004-999
/ CURRENT APPLICATION NUMBER: US/09/792,024
/ NUMBER OF SEQ ID NOS: 490
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 42
/ LENGTH: 1299
/ TYPE: DNA
/ ORGANISM: Candida albicans
US-09-792-024-42
```

```
Query Match 68.6%; Score 891; DB 4; Length 1299;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1291; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```

```
QY 1 ATGTCAAAAGCATTTAGTGCACCTGGAAGGCAATTTCTGCTGATGATTTGGTCTT 60
DB 1 ATGTCAAAAGCATTTAGTGCACCTGGAAGGCAATTTCTGCTGATGATTTGGTCTT 60
QY 61 GAGCCAAATTTAGATGCTTATGTGACAGCATTTGTATCAAGATGATGAGTTATACA 120
DB 61 GAGCCAAATTTAGATGCTTATGTGACAGCATTTGTATCAAGATGATGAGTTATACA 120
QY 121 CCAAAAGGACACGATTGAAAGATCTAGAAATCAAAATTTCTTACCCCAATTTGCAAC 180
DB 121 CCAAAAGGACACGATTGAAAGATCTAGAAATCAAAATTTCTTACCCCAATTTGCAAC 180
QY 181 GGAAGATGGAAATATCAATATCATCAAAATACAGAGAGCCAGAGAGTTCACTACGC 240
DB 181 GGAAGATGGAAATATCAATATCATCAAAATACAGAGAGCCAGAGAGTTCACTACGC 240
QY 241 ATAAATCCATTTTGAAGGCAACTATATTCATCTTTAGTTATTTCAACGACCGAA 300
DB 241 ATAAATCCATTTTGAAGGCAACTATATTCATCTTTAGTTATTTCAACGACCGAA 300
```

```
DB 241 ATAAATCCATTTTGAAGGCAACTATATTCATCTTTAGTTATTTCAACGACCGAA 300
QY 301 GCATTTGATCTTGAATATCATATTTACTCAGACCTCGATATCATTCACAGAAAGATCT 360
DB 301 GCATTTGATCTTGAATATCATATTTACTCAGACCTCGATATCATTCACAGAAAGATCT 360
QY 361 GAAACCAAGACATCTCCGATGAGAGAAAAACATTTCTTATCCATTCCTGTCGATTAAC 420
DB 361 GAAACCAAGACATCTCCGATGAGAGAAAAACATTTCTTATCCATTCCTGTCGATTAAC 420
QY 421 GAAGTGAAGAAAGACCGATTAAGTTGATTCATGCGCAGAGATTAAGTTGTTGCCAAGT 480
DB 421 GAAGTGAAGAAAGACCGATTAAGTTGATTCATGCGCAGAGATTAAGTTGTTGCCAAGT 480
QY 481 TTATATATCCATTTTATCCCAATGTTATCAGTACGATTAAGATATTTTGCACACGTT 540
DB 481 TTATATATCCATTTTATCCCAATGTTATCAGTACGATTAAGATATTTTGCACACGTT 540
QY 541 GCACGATTTGACATTTGATATGCCCAAAAAAGATAGATCTGGGTTGATGTTGCAACT 600
DB 541 GCACGATTTGACATTTGATATGCCCAAAAAAGATAGATCTGGGTTGATGTTGCAACT 600
QY 601 GCAATTTATGCTCTGATTTGATATAGAGATTTCAAGCAGCTTTGATTAATGACGCTTT 660
DB 601 GCAATTTATGCTCTGATTTGATATAGAGATTTCAAGCAGCTTTGATTAATGACGCTTT 660
QY 661 CAGGTTCTAGAAATGATCTCTGAGAAATTTCCCAAGTTGAAAAATTTGATTTGAAAGT 720
DB 661 CAGGTTCTAGAAATGATCTCTGAGAAATTTCCCAAGTTGAAAAATTTGATTTGAAAGT 720
QY 721 AACTGGAAATTCACAAATGAAAGATGATACATTCACATGAGATTCAGATTTATTAAGGT 780
DB 721 AACTGGAAATTCACAAATGAAAGATGATACATTCACATGAGATTCAGATTTATTAAGGT 780
QY 781 GACGTCAAGGGTGTCTGAGAAACACCAATTTGATTCAGATTCATCAATGAGAAAAAG 840
DB 781 GACGTCAAGGGTGTCTGAGAAACACCAATTTGATTCAGATTCATCAATGAGAAAAAG 840
QY 841 GAAAAGCCAGAAAGAGCTCTGTTGTATGACACGCTTAATATGATGCTTATTAACGTTT 900
DB 841 GAAAAGCCAGAAAGAGCTCTGTTGTATGACACGCTTAATATGATGCTTATTAACGTTT 900
QY 901 ATGAAGAAATGAGGAAATGCGTGAATAATGCACTCAGACCCAGAGACTTAATTTAA 960
DB 901 ATGAAGAAATGAGGAAATGCGTGAATAATGCACTCAGACCCAGAGACTTAATTTAA 960
QY 961 GAGTTAGATCATTTCTGTTGAGCCTTTGACTGTTGCGATTAAGACATCAGAAAAGGTTA 1020
DB 961 GAGTTAGATCATTTCTGTTGAGCCTTTGACTGTTGCGATTAAGACATCAGAAAAGGTTA 1020
QY 1021 CAAGCATTAACACAAAAATCAGAGGTTCCATTTGAACCTGATGTCACAAACCAAGTTTG 1080
DB 1021 CAAGCATTAACACAAAAATCAGAGGTTCCATTTGAACCTGATGTCACAAACCAAGTTTG 1080
QY 1081 GACCGTTGCAAGAGATTCCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
DB 1081 GACCGTTGCAAGAGATTCCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
QY 1141 GATGCAATGCTGTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1200
DB 1141 GATGCAATGCTGTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1200
QY 1201 GAAAAATCCAGATTTATTTATATGTTTACTGGTGTGTTGTTGTTGTTGTTGTTGTTGTT 1260
DB 1201 GAAAAATCCAGATTTATTTATATGTTTACTGGTGTGTTGTTGTTGTTGTTGTTGTTGTT 1260
QY 1261 GTACTTGAAGAAAAACCAAGAACTATATAGTTTATTA 1299
DB 1261 GTACTTGAAGAAAAACCAAGAACTATATAGTTTATTA 1299
```

```
RESULT 2
US-09-248-796A-3819
```

```
; Sequence 3819, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3819
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1194)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unkno
US-09-248-796A-3819
```

```
Query Match      44.1%; Score 573; DB 4; Length 1230;
Best Local Similarity 100.0%; Pred. No. 4.8e-276;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 586 TTTGATGTTGCACTGCAATTTATGCTGATTTGATATAGAAAGATTTCAGCCAGCTTTG 645
DB 613 TTTGATGTTGCACTGCAATTTATGCTGATTTGATATAGAAAGATTTCAGCCAGCTTTG 672
QY 646 ATAAATGACGTGTTTCAGGTTCTAGAAAGTGAATCCTGAGAAATGTTCCCAAGAGTTGAA 705
DB 673 ATAAATGACGTGTTTCAGGTTCTAGAAAGTGAATCCTGAGAAATGTTCCCAAGAGTTGAA 732
QY 706 AAATGATTTGAAGTGAATCTGGAATTTCAACAATGAAGATGATACATTAACATACGGAATC 765
DB 723 AAATGATTTGAAGTGAATCTGGAATTTCAACAATGAAGATGATACATTAACATACGGAATC 792
QY 766 AAGTTATTTAATGGGTGACGTCAAGGTTGCTCAAGAAACCCCAATTGGTATCAAGAGTA 825
DB 793 AAGTTATTTAATGGGTGACGTCAAGGTTGCTCAAGAAACCCCAATTGGTATCAAGAGTA 852
QY 826 CTCGAATGGAAGAAAGAAAGCCAGAAAGAGCTCTGTTGATGACCAAGCTTAATAGT 885
DB 853 CTCGAATGGAAGAAAGAAAGCCAGAAAGAGCTCTGTTGATGACCAAGCTTAATAGT 912
QY 886 GCCAATTTACAGTTTATGAAAGAAATTTGAGGAAATGCTGAAATAATACGACTCAGACCCA 945
DB 913 GCCAATTTACAGTTTATGAAAGAAATTTGAGGAAATGCTGAAATAATACGACTCAGACCCA 972
QY 946 GAGACTTATTTAAAGAGTTAATCATCTCTGTTGAGCTTTGACTGTTGCGATTAAAGAC 1005
DB 973 GAGACTTATTTAAAGAGTTAATCATCTCTGTTGAGCTTTGACTGTTGCGATTAAAGAC 1032
QY 1006 ATCAGAAAAGGGTTTCAAGCAATTAACACAAAATCAGAGGTTCCATTAAGACCTGATGTC 1065
DB 1033 ATCAGAAAAGGGTTTCAAGCAATTAACACAAAATCAGAGGTTCCATTAAGACCTGATGTC 1092
QY 1066 CAACCCAGAGTTTGGACCGTTGTCAAGAGATTCCTGTTGTTGTTGTTGTTGTTGTTGCA 1125
DB 1093 CAACCCAGAGTTTGGACCGTTGTCAAGAGATTCCTGTTGTTGTTGTTGTTGTTGTTGCA 1152
QY 1126 GGTGCTGGTGAATACGATCAATAGCTGTATTA 1158
DB 1153 GGTGCTGGTGAATACGATCAATAGCTGTATTA 1185
```

```
RESULT 3
US-09-248-796A-3820
; Sequence 3820, Application US/09248796A
; Patent No. 6747137
```

```
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3820
; LENGTH: 765
; TYPE: DNA
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (762)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unkno
US-09-248-796A-3820
```

```
Query Match      28.2%; Score 366; DB 4; Length 765;
Best Local Similarity 99.1%; Pred. No. 9.5e-173;
Matches 666; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 1 ATGTCAAAAGCATTTAGTACCTGGAAGACATTTCTGCTGGTGGATATTGGTTCTT 60
DB 64 ATGTCAAAAGCATTTAGTACCTGGAAGACATTTCTGCTGGTGGATATTGGTTCTT 123
QY 61 GAGCCAAATTTATGATGCTTATGTTGACAGCATTTGTCATCAAGATCAGTACAGTTATACA 120
DB 124 GAGCCAAATTTATGATGCTTATGTTGACAGCATTTGTCATCAAGATCAGTACAGTTATACA 183
QY 121 CCAAAAAGAACCGATTTGAAAGAAATCTGAAATCAAAATTTCTTCAACCCCAATTTGCAAC 180
DB 184 CCAAAAAGAACCGATTTGAAAGAAATCTGAAATCAAAATTTCTTCAACCCCAATTTGCAAC 243
QY 181 GGAGAAATGGGAATATCATCATATCATCAATACAGAGAAAGCCAGAGAGTTGAGTCACGC 240
DB 244 GGAGAAATGGGAATATCATCATATCATCAATACAGAGAAAGCCAGAGAGTTGAGTCACGC 303
QY 241 ATAAATCCATTTTATAGGCAACTATATATTCATCGTTTATGCTTATATTCACACCGAA 300
DB 304 ATAAATCCATTTTATAGGCAACTATATATTCATCGTTTATGCTTATATTCACACCGAA 363
QY 301 GCATTTGATCTTGAATATCATCTATCTCAGACCTCGATATCATTCACAAGAATACT 360
DB 364 GCATTTGATCTTGAATATCATCTATCTCAGACCTCGATATCATTCACAAGAATACT 423
QY 361 GAAACCAAGACATCTCGAATGAGAAAGAAACATTTCTTATCATTCCTGTCGCAATTAAC 420
DB 424 GAAACCAAGACATCTCGAATGAGAAAGAAACATTTCTTATCATTCCTGTCGCAATTAAC 483
QY 421 GAAGTGAAGAAAGACCGATTAAGTTATCGGACAGATTAAGTCAAGTTGTTGCCACAGT 480
DB 484 GAAGTGAAGAAAGACCGATTAAGTTATCGGACAGATTAAGTCAAGTTGTTGCCACAGT 543
QY 481 TTATATCCCATTTTATCCCAATGTTATCAATGATCAATTAAGATATTTGCAACGTT 540
DB 544 TTATATCCCATTTTATCCCAATGTTATCAATGATCAATTAAGATATTTGCAACGTT 603
QY 541 GCACAGATTTGACATTTGTTATGCCCCAAAAGATAGAGATCGGGTTGATGTTGCAACT 600
DB 604 GCACAGATTTGACATTTGTTATGCCCCAAAAGATAGAGATCGGGTTGATGTTGCAACT 663
QY 601 GCAATTTATGCTGATTTGATATAGAGATTTACGACGATTTGATTAATACGCTGTTT 660
DB 664 GCAATTTATGCTGATTTGATATAGAGATTTACGACGATTTGATTAATACGCTGTTT 723
QY 661 CAGGTTCTAGAA 672
DB 724 CAGGTTCTAGAA 735
```

RESULT 4

US-09-792-024-348/C
; Sequence 348, Application US/09792024
; Patent No. 6783985
; GENERAL INFORMATION:
; APPLICANT: Roemer, Terry
; APPLICANT: Jlang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug
; FILE REFERENCE: 10182-004-999
; CURRENT APPLICATION NUMBER: US/09/792,024
; CURRENT FILING DATE: 2001-02-20
; NUMBER OF SEQ ID NOS: 490
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 348
; LENGTH: 90
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DNA primer
US-09-792-024-348

Query Match
Best Local Similarity 100.0%; Score 65; DB 4; Length 90;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTCAAAGCATTTAGTGCACCGGAAAGCATTTCTGCTGCTGATTTGTTCTT 60
DB 65 ATGTCAAAGCATTTAGTGCACCGGAAAGCATTTCTGCTGCTGATTTGTTCTT 6
QY 61 GAGCC 65
DB 5 GAGCC 1

RESULT 5

US-09-792-024-470/C
; Sequence 470, Application US/09792024
; Patent No. 6783985
; GENERAL INFORMATION:
; APPLICANT: Roemer, Terry
; APPLICANT: Jlang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug
; FILE REFERENCE: 10182-004-999
; CURRENT APPLICATION NUMBER: US/09/792,024
; CURRENT FILING DATE: 2001-02-20
; NUMBER OF SEQ ID NOS: 490
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 470
; LENGTH: 22
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: DNA primer
US-09-792-024-470

Query Match
Best Local Similarity 1.7%; Score 22; DB 4; Length 22;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 GGATTAGTGTCAAGTTGTTGCCA 475
DB 22 GGATTAGTGTCAAGTTGTTGCCA 1

RESULT 6

US-09-910-009A-157

; Sequence 157, Application US/09910009A
; Patent No. 6727226
; GENERAL INFORMATION:
; APPLICANT: University of Utah Research Foundation
; APPLICANT: Cognetix, Inc.
; APPLICANT: Oliviera, Baldomero M.
; APPLICANT: McIntosh, J. Michael
; APPLICANT: Garrett, James E.
; APPLICANT: Watkins, Maren
; APPLICANT: Cruz, Lourdes J.
; APPLICANT: Shon, Ki-Joon
; APPLICANT: Jacobsen, Richard
; APPLICANT: Jones, Robert M.
; APPLICANT: Cartier, G. Edward
; APPLICANT: Shen, Greg S.
; APPLICANT: Wagstaff, John D.
; TITLE OF INVENTION: Mg-Conopeptides
; FILE REFERENCE: 2314-242
; CURRENT APPLICATION NUMBER: US/09/910,009A
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US 60/219,619
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 60/245,157
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: US 60/264,319
; PRIOR FILING DATE: 2001-01-29
; PRIOR APPLICATION NUMBER: US 60/277,270
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 520
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 157
; LENGTH: 355
; TYPE: DNA
; ORGANISM: *Conus marmoreus*
US-09-910-009A-157

Query Match
Best Local Similarity 1.5%; Score 19; DB 4; Length 355;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 861 TGTTGTGTATGACCACTT 879
DB 262 TGTTGTGTATGACCACTT 280

RESULT 7

US-09-248-796A-2118/C
; Sequence 2118, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 2118
; LENGTH: 558
; TYPE: DNA
; ORGANISM: *Candida albicans*
US-09-248-796A-2118

Query Match
Best Local Similarity 1.5%; Score 19; DB 4; Length 558;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1091 AAGAGATTCTGTTGTCT 1109

Db 42 AAGAGATTCCTGTGTGT 24

RESULT 8

US-09-495-050A-89/C
Sequence 89, Application US/09495050A
Patent No. 6492505
GENERAL INFORMATION:
APPLICANT: Roopa, Reddy
APPLICANT: Guegler, Karl, J.
APPLICANT: Au-Young, Janice
TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED F
FILE REFERENCE: PA-0013 US
CURRENT APPLICATION NUMBER: US/09/495,050A
CURRENT FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: 60/118,318
PRIOR FILING DATE: February 1, 1999
NUMBER OF SEQ ID NOS: 305
SOFTWARE: PERL Program
SEQ ID NO 89
LENGTH: 593
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6492505 1330882CT1
US-09-495-050A-89

Query Match 1.5%; Score 19; DB 4; Length 593;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 GAGAAAAACATTTCTTTA 401
Db 432 GAGAAAAACATTTCTTTA 414

RESULT 9

US-08-803-973-3/C
Sequence 3, Application US/08803973
Patent No. 5773701
GENERAL INFORMATION:
APPLICANT: Braun, Carl J.
APPLICANT: Hemenway, Cynthia L.
APPLICANT: Turner, Nilgun E.
TITLE OF INVENTION: Plants Resistant to Infection by PVX
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BBAF
STREET: 700 Chesterfield Village Parkway
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63198
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/803,973
FILING DATE: 21-FEB-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/804,862
FILING DATE: 04-DEC-1991
APPLICATION NUMBER: US 07/771,912
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hoerner Jr., Dennis R.
REGISTRATION NUMBER: 30,914
REFERENCE/DOCKET NUMBER: 38-21(10541)A

TELECOMMUNICATION INFORMATION:
TELEPHONE: (314)537-6099
TELEFAX: (314)537-6047
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-803-973-3

Query Match 1.4%; Score 18; DB 1; Length 84;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1101 TGGTGTGTGTGTGTGT 1118
Db 32 TGGTGTGTGTGTGTGT 15

RESULT 10

US-08-803-972-3/C
Sequence 3, Application US/08803972
Patent No. 5792937
GENERAL INFORMATION:
APPLICANT: Braun, Carl J.
APPLICANT: Hemenway, Cynthia L.
APPLICANT: Turner, Nilgun E.
TITLE OF INVENTION: Plants Resistant to Infection by PVX
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BBAF
STREET: 700 Chesterfield Village Parkway
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63198
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/803,972
FILING DATE: 21-FEB-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/804,862
FILING DATE: 04-DEC-1991
APPLICATION NUMBER: US 07/771,912
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hoerner Jr., Dennis R.
REGISTRATION NUMBER: 30,914
REFERENCE/DOCKET NUMBER: 38-21(10541)A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314)537-6099
TELEFAX: (314)537-6047
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-803-972-3

Query Match 1.4%; Score 18; DB 1; Length 84;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1101 TGGTGTGTGTGTGTGT 1118

Db 32 TGTGTGTGTGTGTGTGT 15

RESULT 11
US-08-956-171E-1674/c
; Sequence 1674, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunach
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 1674:
SEQUENCE CHARACTERISTICS:
LENGTH: 109 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

US-08-956-171E-1674

Query Match 1.4%; Score 18; DB 4; Length 109;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1105 TGTGTGTGTGTGTGT 1122
Db 66 TGTGTGTGTGTGTGT 49

RESULT 12
US-08-781-986A-1674/c
; Sequence 1674, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
; APPLICANT: Charles Kunach
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 1674:
SEQUENCE CHARACTERISTICS:
LENGTH: 109 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

US-08-781-986A-1674

Query Match 1.4%; Score 18; DB 4; Length 109;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1105 TGTGTGTGTGTGTGT 1122
Db 66 TGTGTGTGTGTGTGT 49

RESULT 13
US-08-956-171E-4078
; Sequence 4078, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunach
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996

APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 4078:
SEQUENCE CHARACTERISTICS:
LENGTH: 351 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 4078:
US-08-956-171E-4078

Query Match 1.4%; Score 18; DB 4; Length 351;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 948 GACTTATATTAAGATT 965
Db 62 GACTTATATTAAGATT 79

RESULT 14
US-08-781-986A-4078
Sequence 4078, Application US/08781986A
Patent No. 6737248
GENERAL INFORMATION:
APPLICANT: Charles Kunach
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 4078:
SEQUENCE CHARACTERISTICS:
LENGTH: 351 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-4078

Query Match 1.4%; Score 18; DB 4; Length 351;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 948 GACTTATATTAAGATT 965
Db 62 GACTTATATTAAGATT 79

RESULT 15
US-08-956-171E-3865/C
Sequence 3865, Application US/08956171E
Patent No. 6593114
GENERAL INFORMATION:
APPLICANT: Charles Kunach
Gail H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 3865:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 3865:
US-08-956-171E-3865

Query Match 1.4%; Score 18; DB 4; Length 400;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 948 GACTTATATTAAGATT 965
Db 145 GACTTATATTAAGATT 128

RESULT 16
US-08-956-171E-3933
Sequence 3933, Application US/08956171E
Patent No. 6593114
GENERAL INFORMATION:
APPLICANT: Charles Kunach
Gail H. Choi
Patrick S. Dillon
Craig A. Rosen


```

;
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
;
; INFORMATION FOR SEQ ID NO: 3933:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 3933:
;
; US-08-956-171E-3933
;
; Query Match 1.4%; Score 18; DB 4; Length 400;
; Best Local Similarity 100.0%; Pred. No. 43;
; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 948 GACTTATATTAAAGATT 965
; Db 18 GACTTATATTAAAGATT 35
;
; RESULT 17
; US-08-956-171E-4018
; Sequence 4018, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
;
; US-08-956-171E-4018
```

```

;
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
;
; INFORMATION FOR SEQ ID NO: 4018:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 4018:
;
; US-08-956-171E-4018
;
; Query Match 1.4%; Score 18; DB 4; Length 400;
; Best Local Similarity 100.0%; Pred. No. 43;
; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 948 GACTTATATTAAAGATT 965
; Db 266 GACTTATATTAAAGATT 283
;
; RESULT 18
; US-08-781-986A-3865/c
; Sequence 3865, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8512
; TELEFAX: (301) 309-8512
;
; INFORMATION FOR SEQ ID NO: 3865:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 base pairs
;
; US-08-781-986A-3865/c
```

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-3865

Query Match 1.4%; Score 18; DB 4; Length 400;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 948 GACTTATATTAAAGATT 965
DB 145 GACTTATATTAAAGATT 128

RESULT 19
US-08-781-986A-3933
Sequence 3933; Application US/08781986A
Patent No. 6737248
GENERAL INFORMATION:
APPLICANT: Charles Kunach
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 3933:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-3933

Query Match 1.4%; Score 18; DB 4; Length 400;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 948 GACTTATATTAAAGATT 965
DB 18 GACTTATATTAAAGATT 35

RESULT 20
US-08-781-986A-4018
Sequence 4018; Application US/08781986A
Patent No. 6737248
GENERAL INFORMATION:
APPLICANT: Charles Kunach
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255

CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 4018:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-4018

Query Match 1.4%; Score 18; DB 4; Length 400;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 948 GACTTATATTAAAGATT 965
DB 266 GACTTATATTAAAGATT 283

RESULT 21
US-09-543-681A-3991
Sequence 3991; Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 3991
LENGTH: 1431
TYPE: DNA
ORGANISM: Proteus mirabilis
US-09-543-681A-3991

Query Match 1.4%; Score 18; DB 4; Length 1431;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1241 TCGAAGACCAACAGAG 1258
DB 20 TCGAAGACCAACAGAG 37

RESULT 22

US-09-809-665A-17/c
; Sequence 17, Application US/09809665A
; Patent No. 6790950
; GENERAL INFORMATION:
; APPLICANT: Lowery E., David, et al.
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions
; FILE REFERENCE: 28341/00435
; CURRENT APPLICATION NUMBER: US/09/809,665A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/153,453
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: 60/128,689
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 09/545,199
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 17
; LENGTH: 3247
; TYPE: DNA
; ORGANISM: Pasteurella multocida
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1479)..(3245)
; OTHER INFORMATION: fhax2
US-09-809-665A-17

Query Match 1.4%; Score 18; DB 4; Length 3247;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 669 AGAAGTATCTCTGAGAA 686
Db 1443 AGAAGTATCTCTGAGAA 1426

RESULT 23
US-09-809-665A-19/c
; Sequence 19, Application US/09809665A
; Patent No. 6790950
; GENERAL INFORMATION:
; APPLICANT: Lowery E., David, et al.
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions
; FILE REFERENCE: 28341/00435
; CURRENT APPLICATION NUMBER: US/09/809,665A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/153,453
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: 60/128,689
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 09/545,199
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 19
; LENGTH: 3247
; TYPE: DNA
; ORGANISM: Pasteurella multocida
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1446)
; OTHER INFORMATION: fhac
US-09-809-665A-19

Query Match 1.4%; Score 18; DB 4; Length 3247;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 669 AGAAGTATCTCTGAGAA 686
Db 1443 AGAAGTATCTCTGAGAA 1426

RESULT 24
US-09-620-312D-240/c
; Sequence 240, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yundong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Dymnac, Radolje T.
; TITLE OF INVENTION: No. 6569662e1 Nucleic Acids and
; FILE REFERENCE: 784CIP28
; CURRENT APPLICATION NUMBER: US/09/620,312D
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: PL_FL_genes Version 1.0
; SEQ ID NO 240
; LENGTH: 4953
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(4953)
US-09-620-312D-240

Query Match 1.4%; Score 18; DB 4; Length 4953;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 874 CAGCTTAATGTCCTCAAT 891
Db 4194 CAGCTTAATGTCCTCAAT 4177

RESULT 25
US-08-956-171B-143/c
; Sequence 143, Application US/08956171B
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; APPLICANT: Gil H. Choi
; APPLICANT: Patrick S. Dillon
; APPLICANT: Craig A. Rosen
; APPLICANT: Steven C. Barash
; APPLICANT: Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 143:
SEQUENCE CHARACTERISTICS:
LENGTH: 10320 base pairs
TYPE: nucleic acid
STRADEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 143:
US-08-956-171E-143

Query Match 1.4%; Score 18; DB 4; Length 10320;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 948 GACTTATATTAAAGCTT 965
DB 5759 GACTTATATTAAAGCTT 5742

RESULT 26
US-08-781-986A-143/C
Sequence 143, Application US/08781986A
Patent No. 6737248
GENERAL INFORMATION:
APPLICANT: Charles Kunach
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248BP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 143:
SEQUENCE CHARACTERISTICS:

LENGTH: 10320 base pairs
TYPE: nucleic acid
STRADEDNESS: double
TOPOLOGY: linear
US-08-781-986A-143

Query Match 1.4%; Score 18; DB 4; Length 10320;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 948 GACTTATATTAAAGCTT 965
DB 5759 GACTTATATTAAAGCTT 5742

RESULT 27
US-09-790-988-1/C
Sequence 1, Application US/09790988
Patent No. 6632935
GENERAL INFORMATION:
APPLICANT: SHIGENOBU, SHUJI
APPLICANT: MATANABE, HIDEMI
APPLICANT: HATTORI, MASAHIRA
APPLICANT: SAKAKI, YOSHIYUKI
TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
FILE REFERENCE: 081356/0159
CURRENT APPLICATION NUMBER: US/09/790,988
CURRENT FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: JP2000-107160
PRIOR FILING DATE: 2000-04-07
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 640681
TYPE: DNA
ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 1.4%; Score 18; DB 4; Length 640681;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 AACTATATTCATCGTTT 278
DB 194919 AACTATATTCATCGTTT 194902

RESULT 28
US-08-016-863-12/C
Sequence 12, Application US/08016863
Patent No. 5602003
GENERAL INFORMATION:
APPLICANT: Pierce, J. M.
APPLICANT: Shoreibah, Mohamed G.
APPLICANT: Adler, Beverly
APPLICANT: Fregien, Nevils L.
TITLE OF INVENTION: N-Acetylglucosaminyltransferase V
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/016,863

FILED DATE: 19930210
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/905,795
FILING DATE: 29-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Felder, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 34-92A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)499-8080
TELEFAX: (303)499-8089
TELEX: 823189
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 27 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (other)
HYPOTHEICAL: NO
US-08-016-863-12

Query Match 1.3%; Score 17; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 334 CCTGATATCATTCACA 350
DB 24 CCTGATATCATTCACA 8

RESULT 29
US-08-276-968A-12/C
Sequence 12, Application US/08276968A
Patent No. 6015701
GENERAL INFORMATION:
APPLICANT: Pierce, James M.
APPLICANT: Shoreibah, Mohamed G.
APPLICANT: Adler, Beverly L.
TITLE OF INVENTION: N-Acetylglucosaminyltransferase V
TITLE OF INVENTION: Proteins and Sequences
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,968A
FILING DATE: 19-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/905,795
FILING DATE: 29-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/016,863
FILING DATE: 10-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Felder, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 34-92D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)-499-8080
TELEFAX: (303)-499-8089

TELEX: 49617824
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 27 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
HYPOTHEICAL: NO
ANTI-SENSE: YES
US-08-276-968A-12

Query Match 1.3%; Score 17; DB 3; Length 27;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 334 CCTGATATCATTCACA 350
DB 24 CCTGATATCATTCACA 8

RESULT 30
US-08-956-171E-1400/C
Sequence 1400, Application US/08956171E
Patent No. 6593114
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
Gil H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44M storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 1400:
SEQUENCE CHARACTERISTICS:
LENGTH: 107 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1400:
US-08-956-171E-1400
Query Match 1.3%; Score 17; DB 4; Length 107;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;


```
/ ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2027

Query Match
Best Local Similarity 1.3%; Score 17; DB 3; Length 279;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 393 ATTCTTTACCATCTC 409
Db 161 ATTCTTTACCATCTC 145

RESULT 35
US-09-248-796A-10735
/ Sequence 10735, Application US/09248796A
/ Patent No. 6747137
/ GENERAL INFORMATION:
/ APPLICANT: Keith Weinstock et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
/ FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS
/ CURRENT APPLICATION NUMBER: US/09/248,796A
/ PRIOR FILING DATE: 1999-02-12
/ PRIOR APPLICATION NUMBER: US 60/074,725
/ PRIOR FILING DATE: 1998-02-13
/ PRIOR APPLICATION NUMBER: US 60/096,409
/ NUMBER OF SEQ ID NOS: 28208
/ SEQ ID NO 10735
/ LENGTH: 291
/ TYPE: DNA
/ ORGANISM: Candida albicans
US-09-248-796A-10735

Query Match
Best Local Similarity 1.3%; Score 17; DB 4; Length 291;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 705 AAAATGATTGAAGTA 721
Db 240 AAAATGATTGAAGTA 256

RESULT 36
US-09-540-236-1526/C
/ Sequence 1526, Application US/09540236
/ Patent No. 6673910
/ GENERAL INFORMATION:
/ APPLICANT: Gary L. Breton et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
/ FILE REFERENCE: 2709.2005-001
/ CURRENT APPLICATION NUMBER: US/09/540,236
/ CURRENT FILING DATE: 2000-04-04
/ NUMBER OF SEQ ID NOS: 3840
/ SEQ ID NO 1526
/ LENGTH: 348
/ TYPE: DNA
/ ORGANISM: M.catarrahalls
US-09-540-236-1526

Query Match
Best Local Similarity 1.3%; Score 17; DB 4; Length 348;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 301 GCATTGATCTTGAAT 317
Db 52 GCATTGATCTTGAAT 36

RESULT 37
US-09-270-767-1760/C
/ Sequence 1760, Application US/09270767
/ ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1064

Patent No. 6703491
/ GENERAL INFORMATION:
/ APPLICANT: Homburger et al
/ TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
/ FILE REFERENCE: File Reference: 7326-094
/ CURRENT APPLICATION NUMBER: US/09/270,767
/ CURRENT FILING DATE: 1999-03-17
/ NUMBER OF SEQ ID NOS: 62517
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 1760
/ LENGTH: 404
/ TYPE: DNA
/ ORGANISM: Drosophila melanogaster
US-09-270-767-1760

Query Match
Best Local Similarity 1.3%; Score 17; DB 4; Length 404;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 834 GAAAAGGAAAGCCAG 850
Db 45 GAAAAGGAAAGCCAG 29

RESULT 38
US-09-270-767-17042/C
/ Sequence 17042, Application US/09270767
/ Patent No. 6703491
/ GENERAL INFORMATION:
/ APPLICANT: Homburger et al
/ TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
/ FILE REFERENCE: File Reference: 7326-094
/ CURRENT APPLICATION NUMBER: US/09/270,767
/ CURRENT FILING DATE: 1999-03-17
/ NUMBER OF SEQ ID NOS: 62517
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 17042
/ LENGTH: 404
/ TYPE: DNA
/ ORGANISM: Drosophila melanogaster
US-09-270-767-17042

Query Match
Best Local Similarity 1.3%; Score 17; DB 4; Length 404;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 834 GAAAAGGAAAGCCAG 850
Db 45 GAAAAGGAAAGCCAG 29

RESULT 39
US-09-134-001C-1064
/ Sequence 1064, Application US/09134001C
/ Patent No. 6380370
/ GENERAL INFORMATION:
/ APPLICANT: Lynn Doucette-Stamm et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
/ FILE REFERENCE: GTC-007
/ CURRENT APPLICATION NUMBER: US/09/134,001C
/ CURRENT FILING DATE: 1998-08-13
/ PRIOR APPLICATION NUMBER: US 60/064,964
/ PRIOR FILING DATE: 1997-11-08
/ PRIOR APPLICATION NUMBER: US 60/055,779
/ PRIOR FILING DATE: 1997-08-14
/ NUMBER OF SEQ ID NOS: 5674
/ SEQ ID NO 1064
/ LENGTH: 438
/ TYPE: DNA
/ ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1064
```


Query Match 1.3%; Score 17; DB 3; Length 438;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 956 TTAAGAGTAGATCAT 972
Db 158 TTAAGAGTAGATCAT 174

RESULT 40

US-09-270-767-9605
; Sequence 9605, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-084
; CURRENT APPLICATION NUMBER: US/09/270,767
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 9605
; LENGTH: 455
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-9605

Query Match 1.3%; Score 17; DB 4; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 415 ATTACCGAAGTGAAAA 431
Db 257 ATTACCGAAGTGAAAA 273

Search completed: January 26, 2005, 10:01:59
Job time : 117 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 26, 2005, 05:47:53 ; Search time 552 Seconds
(without alignments)
13521.574 Million cell updates/sec

Title: US-10-069-062-6

Perfect score: 1299
Sequence: 1 atgtcaaaagcatcttagtgc.....aagactatataggtttataa 1299

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published Applications NA.*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1286.2	99.0	1299	US-10-032-585-6023	Sequence 6023, Ap
2	132	10.2	1356	US-09-918-740-48	Sequence 48, Appl
3	132	10.2	1356	US-10-006-909-5	Sequence 5, Appl
4	132	10.2	1356	US-10-369-493-25596	Sequence 25596, A
5	132	10.2	1356	US-10-411-066-5	Sequence 5, Appl
6	132	10.2	1356	US-10-450-941-29	Sequence 29, Appl
7	132	10.2	1356	US-10-835-516-48	Sequence 48, Appl
8	132	10.2	1356	US-10-793-639-513	Sequence 513, App
9	132	10.2	4482	US-10-006-909-9	Sequence 9, Appl
10	132	10.2	4482	US-10-411-066-9	Sequence 9, Appl
11	132	10.2	5051	US-10-006-909-12	Sequence 12, Appl
12	132	10.2	5051	US-10-411-066-12	Sequence 12, Appl

13	132	10.2	5463	US-10-006-909-13	Sequence 13, Appl
14	132	10.2	5963	US-10-411-066-13	Sequence 13, Appl
15	132	10.2	7681	US-09-918-740-61	Sequence 61, Appl
16	132	10.2	7681	US-10-835-516-61	Sequence 61, Appl
17	132	10.2	7693	US-09-918-740-58	Sequence 58, Appl
18	132	10.2	7693	US-10-835-516-58	Sequence 58, Appl
19	132	10.2	7695	US-09-918-740-59	Sequence 59, Appl
20	132	10.2	7695	US-10-835-516-59	Sequence 59, Appl
21	132	10.2	8224	US-09-918-740-62	Sequence 62, Appl
22	132	10.2	8224	US-10-835-516-62	Sequence 62, Appl
23	132	10.2	8235	US-09-918-740-60	Sequence 60, Appl
24	132	10.2	8235	US-10-835-516-60	Sequence 60, Appl
25	132	10.2	8400	US-09-918-740-64	Sequence 64, Appl
26	132	10.2	8400	US-10-835-516-64	Sequence 64, Appl
27	132	10.2	9253	US-10-006-909-7	Sequence 7, Appl
28	132	10.2	9253	US-10-411-066-7	Sequence 7, Appl
29	132	10.2	13917	US-09-918-740-72	Sequence 72, Appl
30	132	10.2	13917	US-10-835-516-72	Sequence 72, Appl
31	132	10.2	14623	US-09-918-740-74	Sequence 74, Appl
32	132	10.2	14623	US-10-835-516-74	Sequence 74, Appl
33	132	10.2	14623	US-10-835-516-74	Sequence 74, Appl
34	132	10.2	14623	US-10-835-516-74	Sequence 74, Appl
35	86.6	6.7	1269	US-10-369-493-25849	Sequence 25849, A
36	67.6	5.2	1306	US-10-320-797-2218	Sequence 2218, Ap
37	66	5.1	1356	US-10-128-714-2406	Sequence 2406, Ap
38	65	5.0	65	US-10-032-585-3023	Sequence 3023, Ap
39	62.6	4.8	1455	US-10-128-714-7406	Sequence 7406, Ap
40	62.6	4.8	1509	US-10-128-714-1406	Sequence 1406, Ap
41	62.6	4.8	1509	US-10-128-714-6406	Sequence 6406, Ap
42	62.6	4.8	3508	US-10-128-714-406	Sequence 406, App
43	62.6	4.8	3509	US-10-128-714-5406	Sequence 5406, Ap
44	46.8	3.6	586	US-10-653-047-1245	Sequence 1245, Ap
45	44.6	3.4	1444	US-10-437-963-96128	Sequence 96128, A

ALIGNMENTS

RESULT 1
US-10-032-585-6023
; Sequence 6023, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6023
; LENGTH: 1299
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-6023

Query Match 99.0%; Score 1286.2; DB 15; Length 1299;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1291; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 ATGTCAAAAGCATTTAGTGCACCTGGAAAGCATTTCTGCTGCGTGGATTTGTTCTT 60
DB 1 ATGTCAAAAGCATTTAGTGCACCTGGAAAGCATTTCTGCTGCGTGGATTTGTTCTT 60
QY 61 GAGCCATTTATGATGCTTATGATGACAGATTTGATCATCGAATGATGATTTATTA 120
DB 61 GAGCCATTTATGATGCTTATGATGACAGATTTGATCATCGAATGATGATTTATTA 120
QY 121 CCAAAAGAACGATTTGAAGATCTAGATCAAAATTTCTTCAACCCCAATTTGCAAC 180
DB 121 CCAAAAGAACGATTTGAAGATCTAGATCAAAATTTCTTCAACCCCAATTTGCAAC 180


```

Db      609 TGAAGCGGGTTGAGTGAAGCGCGGACACATATGATCTATCAGATATAGAGATTCC 668
Qy      636 GCCAGCTTGATTAATGACGTGTTTCAGGTTCTAGAAAGATCCGAGAAGTTCCCGAC 695
Db      669 ACCCGCATTAATCTCTAATTTGCCAGATATT-----GGAAGTCTACTTACGGCGTAA 722
Qy      696 AGAGTTGAAAAAATTGATTGAAAGTAACTGGGAATTCAACATGAAGAATGTACATTACC 755
Db      723 ACTGGGGCATTTGGTTGATGAAGAAGACTGGAAATATTAAGATTAAAGTAAACCATTTACC 782
Qy      756 ATACGGAATCAAGTTATTATAGGTGACGTCAAGGGTGGCTCAGAAACCCCAATTGGT 815
Db      783 TTCGGGATTAACTTTATGATGAGCGCATATTAAAGATGGTTCAGAAACAGTAAACCTGGT 842
Qy      816 ATCACAGACTCTCCATGGAAGAAAGAAAGCAGAAAGAGCTGTGTGTATGACCA 875
Db      843 CCAGAAAGTAAATAAATTGGTATGATTCGCATATGCCAGAAAGCTTGAAATATATACAG 902
Qy      876 GCTTAATAGTCCCAATTACAGTTTATGAAGAAATT 911
Db      903 ACTCATCATGCAAAATTTAGATTTATGATGACT 938

```

RESULT 3

```

US-10-006-909-5
; Sequence 5, Application US/10006909
; Publication No. US20030148479A1
; GENERAL INFORMATION:
; APPLICANT: KEASLING, JAY
; APPLICANT: MARTIN, VINCENT
; APPLICANT: PITERA, DOUGLAS
; APPLICANT: KIM, SEON-MON
; APPLICANT: WITHERS III, SYDNOR T.
; APPLICANT: YOSHIKUNI, YASUO
; APPLICANT: NEMMAN, JACK
; APPLICANT: KHEBENIKOV, ARTEM VALENTINOVICH
; TITLE OF INVENTION: BIOSYNTHESIS OF ISOBENTENYL PYROPHOSPHATE
; FILE REFERENCE: 2000-0007
; CURRENT APPLICATION NUMBER: US/10/006,909
; CURRENT FILING DATE: 2002-04-02
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1356
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Phosphomevalonate kinase nucleotide sequence
US-10-006-909-5

```

```

Query Match      10.2%; Score 133; DB 15; Length 1356;
Best Local Similarity 50.5%; Pred. No. 4.1e-23;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;
Qy      3 GTCGAAGATTTAGTGAACCTGGAAGCAATTTCTTGTGCTGATTTTGGTTCTTGA 62
Db      9 GTTGAGAGCTTCAGAGGCCCGAGGGAAGCGTTACTAGCTGGGAGATTTTAACTTTTAA 68
Qy      63 GCCAATTTATGATGCTTATGTGACAGCATTTGTCATCAGCAATSCATGCAAGTTTAAACACC 122
Db      69 TACAAATATGAAGCATTTGTAGTGGATTTATCGGCAAGATTCAGCTGTAGCCATCC 128
Qy      123 AAAAGAAC-----CAGTTTGAAGAATCTAGAAATCAAAATTTCTTCAACCCCAATTGC 176
Db      129 TTACGGTTTATTCGAAGGCTGATTAAGTTGAAAGTCGCTGTAAGTAAGTAACAAATTAA 188
Qy      177 AAACGGAATGGGAATATCACAATATCAATCAATACAG--AGAACCCAGAAAGTTCA 233
Db      189 AGATGGGAGTGGCTGTACCATTAATAGTCTTAAAGTGGCTTCAATTCGTTTCAATAG 248
Qy      234 GTACGCATTAATCATTTTATAGAGCAACTATATTCATCGTTTATGCTTATATTCACACC 293

```

```

Db      249 CGGATCTAAGAACCCCTTTCATTGAATAAGTTATCGCTAACGTATTTAGCTACTTTAAACC 308
Qy      294 GACCGAAGCATTTGATCTTGAA-----TCATCATTTTCTCAGACCTCGATATCA 344
Db      309 TAAATGACGACACTACTGCATAATGAACCTTGTTCTGTTATGATTTTCTCTGATGATGC 368
Qy      345 TTCACAAAGAAATACTGAACCAAGACATCTCGAATGAGAAATAATCTTCTTACCA 404
Db      369 CTACACTTCTCAGAGAGATAGGTTTACGAACATGTGGCAACAGAAATGAGTTTCA 428
Qy      405 TTCTGCTGCCATTACCGAAGTGAAGACCGGATTAGTTCAATGCGCAGATTAGTGC 464
Db      429 TTCGACAGAAATTGAAGAGTTCCCAAAACAGGCGTGGCTCTCCGCGAGTTTACTCAC 488
Qy      465 AGTTGTTCACAAGTTTATATCCCATTTTATCC-----CATGTTATCAGTAC 515
Db      489 AGTTTAACTTACAGCTTTGGCTCTTTTGTATGACCTGGAATAATATGTACAA 548
Qy      516 GATTAAGATTAATTTGACACAGTTCACAGATTCGACATTTGATGCCCCAAAAAGAT 575
Db      549 ATATAGAGAAATTTATTCATAATTTTACACAAATTTGCTCATTTGCAAGCTCAGGTAAT 608
Qy      576 AGGATCTGGTTGATGTTGCAACTGCAATTATATGCTGATTTATATAGAAATTTCA 635
Db      609 TGAAGCGGGTTGATGTAGCGCGGACACATATGATCTATCAGATATAGAAAGATTCC 668
Qy      636 GCCAGCTTGATTAATGACGTGTTCAAGGTTCTAGAAATGATCCGAGAAAGTTCCCGAC 695
Db      669 ACCCGCATTAATCTCTAATTTGCGAGATATT-----GGAAGTCTACTTACGGCGTAA 722
Qy      696 AGAGTTGAAAAAATTGATTGAAAGTAACTGGGAATTCACATGAAGAATGTACATTACC 755
Db      723 ACTGGCGCATTTGGTTGATGAAGAAAGACTGGAATTTAGATTAAAGTAAACCATTTACC 782
Qy      756 ATACGGAATCAAGTTATTATAGGTGACGTCAAGGGTGGCTCAGAAACCCCAATTGGT 815
Db      783 TTCGGGATTAACTTTATGATGAGCGCATATTAAAGATGGTTCAAGAAACAGTAAACCTGGT 842
Qy      816 ATCACAGACTCTCCATGGAAGAAAGAAAGCAGAAAGAGCTGTGTGTATGACCA 875
Db      843 CCAGAAAGTAAATAAATTGGTATGATTCGCATATGCCAGAAAGCTTGAAATATATACAG 902
Qy      876 GCTTAATAGTCCCAATTTACAGTTTATGAAGAAATT 911
Db      903 ACTCATCATGCAAAATTTAGATTTATGATGACT 938

```

RESULT 4

```

US-10-369-493-25596
; Sequence 25596, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR APPLICATION NUMBER: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 25596
; LENGTH: 1356
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-25596

```

```

Query Match      10.2%; Score 133; DB 15; Length 1356;

```

Best Local Similarity 50.5%; Pred. No. 4,1e-23;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

```
QY 3 GTCAAAAGCATTATGACCTGAGAAAGCATTTCTTGCTGGTGATATTGGTTCTTGA 62
Db 9 GTTGAGAGCCTTCAGTGGCCCGAGGAAAGCGTTACAGCTGGTGATATTGGTTTAA 68
QY 63 GCCAATTTATGATGCTTATGTCAGACATTTGATCAGCAATGATGAGTATATAACACC 122
Db 69 TACAAATATGTAAGCATTTTGATGCGATTTATCGGAAATGATGCTGTGATCCATCC 128
QY 123 AAAAGAAC-----CAGTTGAAAGAAATCTAGAAATCAAAATTTCTTCAACCCCAATTGC 176
Db 129 TTAAGGTTTATGCAAGGCTGATGTAAGTTTGAAGTCGTGCAAAAGTAACAATTAA 188
QY 177 AAACGAGAAATGGGAATATCATATCATCAATACAG---AGAAGCCAGAGAAATTGA 233
Db 189 AGATGGGAGTGGCTGTACATATAGTCTTAAAGTGGCTTCATTCTGTTTCATAG 248
QY 234 GTCAAGCATAAATCCATTTTATGAGGCACTATATTATCGTTTATGCTTATATCAACC 293
Db 249 CGGATTTAAGAACCTTTTATGAAAAAGTTATCGTAAAGTATTTAGCTACTTTAAAC 308
QY 294 GACCGAAGCATTTGATCTTGAAA-----TCATCATTTACTCAGACCTGATATGA 344
Db 309 TAAATGAGAGCATCTACTGCAATATGAAACTTGTTCGTTATGATATTCTCTGATGATGC 368
QY 345 TTCAACAAGAACTACTGAAACCAAGCATCTCTGAAATGAGAAAGAAACATTTCTTTACA 404
Db 369 CTACCAATCTCAGAGAGATAGCGTTTACCAACATCTGTGCAACAGAAATGAGTTTCA 428
QY 405 TTCTGTCGCATTACCGAAGTGAAAGAACCGGATAGGTTCACTGGCAGATTTAGTGTGC 464
Db 429 TTTCGACAGAAATTTGAAGAAATTTCCCAAAACAGGGCTGGGCTCTCGGCAAGTTTATGC 488
QY 465 AGTTGTGGCAAGTTTATTTATCCCATTTTATCCC-----CAATGTTATCAGTAC 515
Db 489 AGTTTAACTACAGCTTTGGCCCTCTTTTGTATGAGCATGGAATAATATGATACAA 548
QY 516 GAATTAAGATTTTTCACACAGCTTTCACAGATTTGCACTTTGATGCCCAAAAAAGAT 575
Db 549 ATATAGAGAAATTTATCATATTTTACACAAAGTTGTCTCATTTGCAAGCTCAGGGTAAAT 608
QY 576 AGGATCTGGGTTTATGATGTCACATCTGCAATTTATGCTGATGTATATGAAATTTCA 635
Db 609 TGGAGCGGTTTATGATGTCAGCGGCGAGCATATGATCTATCAGATATGAAATTTCC 668
QY 636 GCCAGCTTTGATTAATGACGTGTTTCAGGTTCTAGAAAGTATCTGAGAAAGTTCCAC 695
Db 669 ACCCGATTAACTCTAATTGGCCAGATAT-----GGAAGTCTACTTAACGAGTAA 722
QY 696 AAGATTGAAAAAATTTGATGAAAGTAACTGGGAATTTCAACATGAAAGATGATATACC 755
Db 723 ACTGGGCGATTTGGTGTATGAAAGAAAGCTGGAATTTTACGATTTAAAGTAAACATTTACC 782
QY 756 ATACGAATCAAGTTATTTATATGAGTACGTCAGAGGTCCTCAGAAACCCCAATTTGT 815
Db 783 TTGGGATTTAATCTTTATGATGTCGCGATATTAAGATGTTTCAGAAACAGTAAACTGG 842
QY 816 ATCAGAGTAACTCAGATGAAAAAGAAAAAGCCAGAAAGAAAGCTCTGTTGTATGACCA 875
Db 843 CCAGAGATTAATAAATTTGATATGATTCGATATGCCAGAAAGCTTGAAATATATACAG 902
QY 876 GCTTAATAGTCCCAATTTTACAGTTTATGAAAGAAAT 911
Db 903 ACTGATCATGCAAAATTTGATTTATGATGATGACT 938
```

RESULT 5
US-10-411-066-5
; Sequence 5, Application US/10411066
; Publication No. US20040005678A1
; GENERAL INFORMATION:

```
; APPLICANT: KEASLING, JAY  
; APPLICANT: MARTIN, VINCENT  
; APPLICANT: PITERA, DOUGLAS  
; APPLICANT: MITTERS III, SYDOR T.  
; APPLICANT: NEMAN, JACK  
; TITLE OF INVENTION: BIOSYNTHESIS OF AMORPHA-4,11-DIENE  
; FILE REFERENCE: 2000-0007.20  
; CURRENT APPLICATION NUMBER: US/10/411,066  
; CURRENT FILING DATE: 2000-04-09  
; PRIOR APPLICATION NUMBER: 10/006,909  
; PRIOR FILING DATE: 2001-12-06  
; NUMBER OF SEQ ID NOS: 124  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 1356  
; TYPE: DNA  
; ORGANISM: Saccharomyces cerevisiae  
US-10-411-066-5
```

Query Match 10.2%; Score 132; DB 16; Length 1356;

Best Local Similarity 50.5%; Pred. No. 4,1e-23;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

```
QY 3 GTCAAAAGCATTATGACCTGAGAAAGCATTTCTTGCTGGTGATATTGGTTCTTGA 62
Db 9 GTTGAGAGCCTTCAGTGGCCCGAGGAAAGCGTTACTAGCTGGTGATATTGGTTTAA 68
QY 63 GCCAATTTATGATGCTTATGTCAGACATTTGATCAGCAATGATGAGTATATAACACC 122
Db 69 TACAAATATGTAAGCATTTTGATGCGATTTATCGGAAATGATGCTGTGATCCATCC 128
QY 123 AAAAGAAC-----CAGTTGAAAGAAATCTAGAAATCAAAATTTCTTCAACCCCAATTGC 176
Db 129 TTAAGGTTTATGCAAGGCTGATGTAAGTTTGAAGTCGTGCAAAAGTAACAATTAA 188
QY 177 AAACGAGAAATGGGAATATCATATCATCAATACAG---AGAAGCCAGAGAAATTGA 233
Db 189 AGATGGGAGTGGCTGTACATATAGTCTTAAAGTGGCTTCATTCTGTTTCATAG 248
QY 234 GTCAAGCATAAATCCATTTTATGAGGCACTATATTATCGTTTATGCTTATATCAACC 293
Db 249 CGGATTTAAGAACCTTTTATGAAAAAGTTATCGTAAAGTATTTAGCTACTTTAAAC 308
QY 294 GACCGAAGCATTTGATCTTGAAA-----TCATCATTTACTCAGACCTGATATGA 344
Db 309 TAAATGAGAGCATCTACTGCAATATGAAACTTGTTCGTTATGATTTCTCTGATGATGC 368
QY 345 TTCAACAAGAACTACTGAAACCAAGCATCTCTGAAATGAGAAAGAAACATTTCTTTACA 404
Db 369 CTACCAATCTCAGAGAGATAGCGTTTACCAACATCTGTGCAACAGAAATGAGTTTCA 428
QY 405 TTCTGTCGCATTACCGAAGTGAAAGAACCGGATAGGTTCACTGGCAGATTTAGTGTGC 464
Db 429 TTTCGACAGAAATTTGAAGAAATTTCCCAAAACAGGGCTGGGCTCTCGGCAAGTTTATGC 488
QY 465 AGTTGTGGCAAGTTTATTTATCCCATTTTATCCC-----CAATGTTATCAGTAC 515
Db 489 AGTTTAACTACAGCTTTGGCCCTCTTTTGTATGAGCATGGAATAATATGATACAA 548
QY 516 GAATTAAGATTTTTCACACAGCTTTCACAGATTTGCACTTTGATGCCCAAAAAAGAT 575
Db 549 ATATAGAGAAATTTATCATATTTTACACAAAGTGTCTCATTTGCAAGCTCAGGGTAAAT 608
QY 576 AGGATCTGGGTTTATGATGTCACATCTGCAATTTATGCTGATGTATATGAAATTTCA 635
Db 609 TGGAGCGGTTTATGATGTCAGCGGCGAGCATATGATCTATCAGATATGAAATTTCC 668
QY 636 GCCAGCTTTGATTAATGACGTGTTTCAGGTTCTAGAAAGTATCTGAGAAAGTTCCAC 695
Db 669 ACCCGATTAACTCTAATTGGCCAGATAT-----GGAAGTCTACTTAACGAGTAA 722
QY 696 AAGATTGAAAAAATTTGATGAAAGTAACTGGGAATTTCAACATGAAAGATGATATACC 755
```

Db 723 ACTGGCGATTGGTTGATGAGAAAGACTGGAAATATTACGATTTAAAGTAAACCATTTACC 782
Qy 756 ATACGAATCAAGTATTATTAATGAGTGAAGGCTGCAAGGCTGCGAAGAACCCAAATTTGGT 815
Db 783 TTCGGGATTAACCTTATGATGAGGCGATTTAAGAAATGGTTCCGAAACAGTAAACCTGGT 842
Qy 816 ATCAGAGTACTCCAAATGAGAAAAAGGAAAGCCAGAAAGAAAGCTCTGTTGTGTATGACCA 875
Db 843 CCAGAAAGTAAATAATTTGGTATGATTCGATATGCGAGAAAGCTTGAATAATATATACAGA 902
Qy 876 GCTTAATAGTGGCAATTTACAGTTTATGAAGAATT 911
Db 903 ACTGATCATGCAAAATTTAGATTATGATGATGACT 938

RESULT 6

US-10-450-941-29
; Sequence 29, Application US/10450941
; Publication No. US20040063182A1
; GENERAL INFORMATION:
; APPLICANT: TOYOTA JIDOSHA KABUSHIKI KAISHA
; TITLE OF INVENTION: A METHOD OF PRODUCING PRENITALCOHOL
; FILE REFERENCE: PH-1444PCT
; CURRENT APPLICATION NUMBER: US/10/450,941
; PRIOR FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: JP2000-401701
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: JP2000-403067
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: JP2001-282978
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
; LENGTH: 1356
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-10-450-941-29

Query Match 10.2%; Score 132; DB 16; Length 1356;
Best Local Similarity 50.5%; Pred. No. 4.1e-23;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;
Qy 3 GTCAAAAGCATTTAGTGAACCTGGAAGGATTTCTTGCTGGTGAATTTGGTTCTTGA 62
Db 9 GTTGAGAGCTTCAAGTCCCGGAGGAAAGCGTTACTAGCTGGTGAATTTAGTTTAA 68
Qy 63 GCCAATTTATGATGCTTATGTCAGACATTTGTCATCAGAAATGTCATGATTAACACC 122
Db 69 TACAAATATGAGCATTTGTAGTCGATTTATCGGCAAGAAATGATGCTGTACCATCC 128
Qy 123 AAAAGAAC-----CAGTTGAAAAGATCTAGATCAAAATTTCTTCAACCCCAATTTGC 176
Db 129 TTAAGGTTTATGCAAGGCTGATTAAGTTGAAGCGCTGTGAAGTAAAGTAAACATTTTA 188
Qy 177 AAACGAGATGGGAATATCATATCATCAATAACAG---AGAAAGCCAGAGATTTCA 233
Db 189 AGATGGAGGAGTGGCTTACATATTAAGTCTTAAGAGTGGCTTATCTGTTTCAATAG 248
Qy 234 GTCAAGCATTAATTCATTTTGAAGGCAATATATTCATGCTTTTATGCTTAATTCAC 293
Db 249 CGGATCTAAGAACCTTTCAATTTGAATAAGTATCGTAAAGTATTTAGCTACTTAAAC 308
Qy 294 GACCGAAGCATTTGATCTTGAA-----TCATCATTTTACTCAGACCTCGGATATCA 344
Db 309 TAACATGAGCACTACTGCAATTAAGAACTTGTTCGTTATTAATTTTCTTGATGATGC 368
Qy 345 TTCACAAGAGATTAATGAAACCAAGCATCTCGAATGAGAAAGAAACATTTCTTTACCA 404
Db 369 CTACCATTTCTCAGAGAGATAGCTTACCAACATCTGTGCAAGAGATTTGATTTTCA 428
Qy 405 TTCTGTCGCAATTACGAAGTGAAGAAACCGGATTAAGTTTCATCGGACAGATTAATGTC 464

Db 429 TTCGACAGAAATTAAGAAAGTTCCAAAACAGGGCTGGGCTCTCTCGGACGTTTACTAC 488
Qy 465 AGTTGTGCGACAAGTTTATTAATCCCATTTTATCC-----CAATGTTATCAGTAC 515
Db 489 AGTTTAACTACAGCTTTGGCCTCTTTTGTATCGACCTGGAATAATATGATACAA 548
Qy 516 GAATTAAGATTTTTCACACAGCTTGACAGATTCGACATTTGATGCCCCAAAAAGAT 575
Db 549 ATATAGAGAAATTAATTAATTTAGACAAAGTTGCTCATTTGCAAGCTCAGGGTAAAT 608
Qy 576 AGGATCTGGGTTTGAATGTCGAATCTGCAATTTATGCTGATTTATTAAGAAATTTCA 635
Db 609 TGAAGCGGGTTTGAATGAGCGCGGACCAATGATCTATCAATTAAGAAAGATTTCC 668
Qy 636 GCAGCTTGATTAATGACGTGTTTCAGGTTCTAGAAATGATCTCGAAGATTTCCCA 695
Db 669 ACCCGCATTAATCTTAATTTGCCAGATTT-----GGAAGTCTACTTAACGCGATTA 722
Qy 696 AGAGTTGAAAAAATGATTAAGTAAGTAACTGGAAATTCAAACATGAAGATTAATTAAC 755
Db 723 ACTGGCGATTGGTTGATGAGAAAGACTGGAATATTAAGATTAACCAATTTACC 782
Qy 756 ATACGAATCAAGTATTATTAATGAGTGAAGTGAAGGCTGCAAGGCTGCAAGAACCCAAATTTGGT 815
Db 783 TTCGGATTAATCTTATGATGATGAGGCGATTTAAGATGTTTCAAGAAACAGTAAACCTGGT 842
Qy 816 ATCAGAGTACTCCAAATGAGAAAAAGGAAAGCCAGAAAGAAAGCTCTGTTGTATGACCA 875
Db 843 CCAGAAAGTAAATAATTTGGTATGATTCGATATGCGCATATGCCAAAGCTTGAATAATATACAGA 902
Qy 876 GCTTAATAGTGGCAATTTACAGTTTATGAAGAATT 911
Db 903 ACTGATCATGCAAAATTTAGATTATGATGATGACT 938

RESULT 7

US-10-835-516-48
; Sequence 48, Application US/10835516
; Publication No. US20040194162A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Frederick
; APPLICANT: Kuehler, Adelheid
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
; FILE REFERENCE: KAS-103XCI
; CURRENT APPLICATION NUMBER: US/10/835,516
; PRIOR FILING DATE: 2004-04-28
; PRIOR APPLICATION NUMBER: US/09/918,740
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/221,703
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 48
; LENGTH: 1356
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-10-835-516-48

Query Match 10.2%; Score 132; DB 18; Length 1356;
Best Local Similarity 50.5%; Pred. No. 4.1e-23;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;
Qy 3 GTCAAAAGCATTTAGTGAACCTGGAAGGATTTCTTGCTGGTGAATTTGGTTCTTGA 62
Db 9 GTTGAGAGCTTCAAGTCCCGGAGGAAAGCGTTACTAGCTGGTGAATTTAGTTTAA 68
Qy 63 GCCAATTTATGATGCTTATGTCAGACATTTGTCATCAGAAATGTCATGATTAACACC 122
Db 69 TACAAATATGAGCATTTGTAGTCGATTTATCGGCAAGAAATGATGCTGTAGCCATCC 128
Qy 123 AAAAGAAC-----CAGTTGAAAAGATCTAGATCAAAATTTCTTCAACCCCAATTTGC 176


```

Db      129 TTAAGGTTTCAGCAAGGTCGTGATTAAGTTGAAGTCGTGTGAAAAAGTAAACAATTAA 188
QY      177 AAACGAGAAATGGGAATATCATATCATCAAAATACG---AGAGCCGAGAGAGTTCA 233
Db      189 AGATGGGAGTGGCTGTACATATTAAGTCTTAAGAGGCTTCCTCTGTTGAGAG 248
QY      234 GTCAAGCAATATTCATTTTAAAGCACTATATTCATGCTTTTAACTTATATTCAC 293
Db      249 CGGATCTAAGAACCTTTTCAATGAAAAAGTATCGTAAGCTATTAAGTACTTTAAAC 308
QY      294 GACCGAAGCATTTGATCTTAAA-----TCATCTTACTCAGACCTTGATATCA 344
Db      309 TAAACATGACGACTACTGCAATGAACCTTGCTGTTATGATATTTCTCTGATATGC 368
QY      345 TTCAACAGAAATATCTGAACCAAGACATCTCGAAATGAGAAAAAATTTCTTTACCA 404
Db      369 CTACCATCTCAGAGAGATAGCGTTACCGAACATCGTGGCAGACAGAAATTTGATTCA 428
QY      405 TTCTCGTCCATTACCGAATGGAAGAACCGGATTAGTTCACTGGCAGATTAAGTGC 464
Db      429 TTGGCAGAGAAATGAAGAAAGTCCCAAAACAGGGCTGGGCTCTCGGCAAGTTTAACTC 488
QY      465 AGTTGTTGCCAAGTTTATATCCATTATCC-----CAATGTTATCAGTAC 515
Db      489 AGTTTAACTACAGCTTTGGCTCTTTTGTATGGAACCTGAAAAATATATGACAA 548
QY      516 GAATTAAGATATTTTGCACACGTTGCACAGATTTGCACATTTTATGCCAAAAAAGAT 575
Db      549 ATATAGAAAGTATTCATATTAATTAAGCAAGTTGCTCATTTGCAAGCTCAGGTTAAAT 608
QY      576 AGATCTGGGTTTGAATGTGCACTGCAATTATATGCTGATTTATATTAAGAAATTTCA 635
Db      609 TGAACGCGGTTTGAATGTGACCGCGCAGCATATGATCATATATGAAGATTTCC 668
QY      636 GCCAGTTTGAATGAAGTGAAGTTCAGGTTTGAAGAAATGCTGGAAGATTTCCCAAC 695
Db      669 ACCCGATTAATCTCTAATTTGCCAATTT-----GAAAGTCTACTTAACGCGTAA 722
QY      696 AGAGTTGAAAAAATGATTAAGAAAGTAACTGGGAATTCAAACATGAAGAATGTACATTACC 755
Db      723 ACTGGCGCATTTGGTATGATGAAGAAAGTGAATATTAAGATTAAGAACATTATTAAC 782
QY      756 ATACGAATCAAGTTTATATGAGGTGACGTCAGAGGTGCTGAGAAACCCAAATTTGT 815
Db      783 TTGGGATTAATCTTAATGATGAGCGCATATTAAGATGTGTCAAGAAACGTAACCTGCT 842
QY      816 ATCAACGACTCTCCATGGAAGAAAGCAAGAACGCTGTTGTATGACCA 875
Db      843 CCAAGAGTAAAGAAATTTGGTATGATTCGCAATATGCCAAGAGCTGAAAAATATATACGA 902
QY      876 GCTTAATAGTCCCAATTTACAGTTTATGAAGAAAT 911
Db      903 ACTGATCATGCAAAATTTAGATTTATGATGAGTACT 938

```

```

RESULT 8
US-10-793-639-513
; Sequence 513, Application US/10793639
; Publication No. US20040199940A1
; GENERAL INFORMATION:
; APPLICANT: Karunananda, Balasubjini
; APPLICANT: Yu, Jaehyuk
; APPLICANT: Kishore, Ganesh M.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED
; FILE REFERENCE: 16516.075
; CURRENT APPLICATION NUMBER: US/10/793,639
; PRIOR FILING DATE: 2004-03-05
; PRIOR APPLICATION NUMBER: US/09/614,221A
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: US 60/142,981
; NUMBER OF SEQ ID NOS: 626

```

```

; SEQ ID NO 513
; LENGTH: 1356
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-10-793-639-513

Query Match      10.2%; Score 132; DB 18; Length 1356;
Best local similarity 50.5%; Pred. No. 4,1e-23;
Matches 4/3; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY      3  GTCAAAAGCATTTAGTGAACCTGAAAAAGCAATTTCTGCTGTGTGATATTTGGTCTTGA 62
Db      9  GTTGAAGCCTTCACTGCCCCAGGAAAGCGTTACTAGCTGTGATATTTAGTTTGA 68
QY      63  GCCAATTAATGATGTTATGACAGCAATTTGTCATCAGAAATGCAATGCAATTAACACC 122
Db      69  TACAAATATGAAAGCATTTGATGAGATTAATCGCAAGAAATGCAATGCTGTAGCCCATCC 128
QY      123  AAAAGAAC-----CAGTTGAAGAATCTAGAAATCAAAATTTCTTACCCCAATTTGC 176
Db      129  TTAAGGTTTCATGCAAGGCTCTGATTAAGTTGAAAGTGGCTGAAAAAGTAAACAATTAA 188
QY      177  AAACGAGAAATGGGAATATCATATCATCAAAATACG---AGAGCCGAGAGAGTTCA 233
Db      189  AGATGGGAGTGGCTGTACCAATATTAAGTCTTAAAGTGGCTTCAATTCCTGTTGATAG 248
QY      234  GTCAAGCAATATTCATTTTAAAGCACTATATTCATGCTTTTAACTTATATTCAC 293
Db      249  CGGATCTAAGAACCTTTTCAATGAAAAAGTATGCTAAGCTATTTAGCTTTAAAC 308
QY      294  GACCGAAGCATTTGATCTTAAA-----TCATCTTACTCAGACCTTGATATCA 344
Db      309  TAAACATGACGACTACTGCAATGAACCTTGCTGTTATGATATTTCTCTGATATGC 368
QY      345  TTCAACAGAAATATCTGAACCAAGACATCTCGAAATGAGAAAAAATTTCTTTACCA 404
Db      369  CTACCATCTCAGAGAGTATCGTTACCGAACATCTGTGGAACAGAAATGAGTTTTCA 428
QY      405  TTCTGTCGCTAATCCGAAATGGAAGAACCGGATTTGCTCATCGGAGATTAAGTGC 464
Db      429  TTGGCAGCAATTTGAAGAAATTTCCAAACAGGCTGGGCTCTCGGCAAGTTTAACTC 488
QY      465  AGTTGTTGCCAAGTTTATATCCATTATCC-----CAATGTTATCAGTAC 515
Db      489  AGTTTAACTACAGCTTTGGCTCTTTTGTATGGAACCTGAAAAATATATGACAA 548
QY      516  GAATTAAGATATTTTGCACACGTTGCACAGATTTGCTATGTTATGCCAAAAAAGAT 575
Db      549  ATATAGAAAGTATTCATATTAATTAAGCAAGATGCTCATTTGTCAAGCTCAGGTTAAAT 608
QY      576  AGATCTGGGTTTGAATGTGCACTGCAATTATATGCTGATTTATGAAGAAATTTCA 635
Db      609  TGAACGCGGTTTGAATGTGACCGCGCAGCATATGATCTATGCAATATGAAGATTTCC 668
QY      636  GCCAGTTTGAATGAAGACGTTTCAAGTTCTAGAAAGTATCCTGAGAAGTTCCCAAC 695
Db      669  ACCCGATTAATCTCTAATTTGCCAATTT-----GAAAGTCTACTTAACGCGTAA 722
QY      696  AGAGTTGAAAAAATGATTAAGAAAGTAACTGGGAATTCAAACATGAAGAATGTACATTACC 755
Db      723  ACTGGCGATTTGGTATGATGAGAGAAAGCTGGAATATTAAGATTAAGAAATTAACCAATTTACC 782
QY      756  ATACGAATCAAGTTTATATGAGGTGACGTCAGAGGTGCTGAGAAACCCAAATTTGT 815
Db      783  TTGGGATTAATCTTAATGATGAGCGCATATTAAGATGTGTCCAGAAACAGTAAATCTGCT 842
QY      816  ATCAACGACTCTCCATGGAAGAAAGCAAGAACGCTGTTGTATGACCA 875
Db      843  CCAAGAGTAAAGAAATTTGGTATGATTCGCAATATGCCAAGAGCTGAAAAATATATACGA 902
QY      876  GCTTAATAGTCCCAATTTACAGTTTATGAAGAAAT 911
Db      903  ACTGATCATGCAAAATTTAGATTTATGATGAGTACT 938

```

RESULT 9
US-10-006-909-9
Sequence 9, Application US/10006909
Publication No. US20030148479A1
GENERAL INFORMATION:
APPLICANT: KEASLING, JAY
APPLICANT: MARTIN, VINCENT
APPLICANT: PITERA, DOUGLAS
APPLICANT: KIM, SEON-WON
APPLICANT: WITHERS III, SYDNOR T.
APPLICANT: YOSHIKUNI, YASUO
APPLICANT: NEWMAN, JACK
APPLICANT: KHEBENIKOV, ARTEM VALENTINOVICH
TITLE OF INVENTION: BIOSYNTHESIS OF ISOENTENYL PYROPHOSPHATE
FILE REFERENCE: 2000-0007
CURRENT APPLICATION NUMBER: US/10/006, 909
CURRENT FILING DATE: 2002-04-02
NUMBER OF SEQ ID NOS: 13
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 9
LENGTH: 4482
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-006-909-9

Query Match 10.2%; Score 132; DB 15; Length 4482;
Best Local Similarity 50.5%; Pred. No. 7.8e-23;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCAAAAGCATTTAGTGCACCTGGAAAAGCATTTCTGCTGGTGGATATTTGGTTCTTGA 62
DB 1606 GTTGAGAGCTTCACAGTCCCGGAGAAAGCGTACAGCTGGGATATTTAGTTTAA 1665
QY 63 GCCAATTTAGATGCTTATGTGACAGCATTTGTCATGACGAATGACGAGTTAAACACC 122
DB 1666 TACAAATATGAGCATTTGTGATCGAATTTATGCGAAGATGATGCTGTACCATCC 1725
QY 123 AAAAGGAC-----CAGTTTGAAGATCTAGAAATCAAAATTTCTTCAACCCCAATTTGC 176
DB 1726 TTAAGGTTTATGCAAGGGTCTGATTAAGTTGAAGCGGTGAAAGTAAACATTTAA 1785
QY 177 AAACGAGATGGGAATATCACTATCAATCAATACAG--AGAAGCCAGAGAATTGA 233
DB 1786 AGATGGGAGTGGCTGTACCATTAAGTCTTAAAGTGGCTTCATTCGTTCGATGAG 1845
QY 234 GTACAGCATTAATCCATTTTATGAGGCACTATATTCATCGTTTATGCTTATTAAC 293
DB 1846 CGATCTTAAGAACCTTTTCAATGAAAAGTTATCGCTAACGATTAAGTACTTTAAAC 1905
QY 294 GACCGAAGCATTTGATCTTGA--TCATCATTTACTAGACCTCGATATCA 344
DB 1906 TAAAGATGACGACTACTGCAATGAAACTGTTCGTTATGATTTTCTGTATGATGC 1965
QY 345 TTCAACAAGAAATATCTGAAACCAAGACATCTCGAATGAGAAAACATTTCTTAC 404
DB 1966 CTACACCTCTCGAGAGATGAGCTTACCGAACATCTGTGCAACAGAAAGATTGATT 2025
QY 405 TTCTGTGTCATTAACGAGTGGAAAAGCCGATTAAGTTCACTGGCGAGATTAGTGC 464
DB 2026 TTTCGACAGATTAAGAAAGTTCCCAAAACAGGCTGGGCTCTCGCGAGTTTATG 2085
QY 465 AGTTGTGTCACAAAGTTTATATCCATTTATCCC-----CAATGTTATCAATAC 515
DB 2086 AGTTTAACTACAGCTTTGGCTCTTTTGTATGAGACCTGGAATAATATGACAA 2145
QY 516 GAATTAAGATTTTGTGACAAAGTTGACAGATTTGACATTTGATGCCCCAAAAAGAT 575
DB 2146 ATATAGAGAAGTTATTCATTAATTTAGACAAAGTTGCTCATTTCTCAAGCTCAG 2205

QY 576 AGATCTGGGTTTGTATGTTGCAACTGCAATTTATGCTGTATTTATAGAAATTTCA 635
DB 2206 TCGAAGCGGTTTGTATGTCGCGGACGACATATGATCTATCAGATATAGAAATTTCC 2265
QY 636 GCCAGCTTTGATTAATGACGTGTTTCAGTCTTACGAAATGATCTCGAAGAGTTCC 695
DB 2266 ACCCGATTAATCTCTTAATTTTCCAGATATTT-----GGAAAGTCTACTACGCG 2319
QY 696 AGAGTTGAAAATATGATTAAGAAAGTAACTGGGAATTCACAACTGAAGATGTATAC 755
DB 2320 ACTGGGCGCTTTGTTGATGAAAGACCTGGAATATTAAGATTAAGATACATTTAC 2379
QY 756 ATACGGAATCAATTATTAATGAGTGCATGCAAGGGTGCAGAAACCCCAATTTG 815
DB 2380 TTGGGATTAATCTTATGATGAGGCGATATTAAGAAATGTTGAGAAACAGTAAAC 2439
QY 816 ATCAGAGTACTCCAAATGGAAGAAAGCAAGAAAGCTCTGTTGTATACCA 875
DB 2440 CCAGAAAGTAAATAATTTGATGATTCGCATATGCCAGAAAGCTTGAAATATAT 2499
QY 876 GCTTAATAGTGCATTTACAGTTTATGAAGAAAT 911
DB 2500 ACTGATCATGCAAAATTTAGATTTATGATGACT 2535

RESULT 10
US-10-411-066-9
Sequence 9, Application US/10411066
Publication No. US20040005678A1
GENERAL INFORMATION:
APPLICANT: KEASLING, JAY
APPLICANT: MARTIN, VINCENT
APPLICANT: PITERA, DOUGLAS
APPLICANT: WITHERS III, SYDNOR T.
APPLICANT: NEWMAN, JACK
TITLE OF INVENTION: BIOSYNTHESIS OF AMORPHA-4,11-DIENE
FILE REFERENCE: 2000-0007.20
CURRENT APPLICATION NUMBER: US/10/411, 066
CURRENT FILING DATE: 2000-04-09
PRIOR APPLICATION NUMBER: 10/006, 909
PRIOR FILING DATE: 2001-12-06
NUMBER OF SEQ ID NOS: 124
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 9
LENGTH: 4482
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-411-066-9

Query Match 10.2%; Score 132; DB 16; Length 4482;
Best Local Similarity 50.5%; Pred. No. 7.8e-23;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCAAAAGCATTTAGTGCACCTGGAAAAGCATTTCTGCTGGTGGATATTTGGTTCTTGA 62
DB 1606 GTTGAGAGCTTCACAGTCCCGGAGAAAGCGTACAGCTGGGATATTTAGTTTAA 1665
QY 63 GCCAATTTAGATGCTTATGTGACAGCATTTGTCATGACGAATGACGAGTTAAACACC 122
DB 1666 TACAAATATGAGCATTTGTGATCGAATTTATGCGAAGATGATGCTGTACCATCC 1725
QY 123 AAAAGGAC-----CAGTTTGAAGATCTAGAAATCAAAATTTCTTCAACCCCAATTTGC 176
DB 1726 TTAAGGTTTATGCAAGGGTCTGATTAAGTTGAAGCGGTGAAAGTAAACATTTAA 1785
QY 177 AAACGAGATGGGAATATCACTATCAATCAATACAG--AGAAGCCAGAGAATTGA 233
DB 1786 AGATGGGAGTGGCTGTACCATTAAGTCTTAAAGTGGCTTCATTCGTTCGATGAG 1845

OY	234	GTTCAGCAGCAATTAATCCATTTTGTAGAGCAACTATATATCAATCGTTTATAGCTATATATCAAC	293
Db	1846	CGGATCTTAAAGCCCTTTCATTTAGAAAAAGTATCCGTAAACGATTTAGCTACTTTTAAACC	1905
OY	234	GAACGAAAGCATTTGAATCTTGAAA-----TCATCATTTTATCTCAGACCCTCGAATATCA	344
Db	1906	TAACTAGACGCACTACGTCAATATAGAAACTTGTCGTTATATGTAATTTCTCTGATGATGCG	1965
OY	345	TTCAACAAGAAAGTACTGAAACCAACATCCCTCGAATGTGAGAAAAAATTTCTTTTACCA	404
Db	1966	CTACCAATTTCTCAGAGAGATAGGTTTACCGAACATCGTGCACACAAAGATTTGAGTTTCA	2023
OY	405	TTCTCGTGCATTACCGAAGTGGAAAAGCCCGATTTAGGTTCATGCGCAGAGTTAGTGTG	464
Db	2026	TTCCGACAGAAATTTGAAAGAAAGTTCCCAAAAGAGGCGTGCGGCTCTCGCAGAGTTTGTAC	2085
OY	465	AGTTGTGCCACAAGTTTATATATCCATTTTATCCC-----CAATGTTATCAGTAC	515
Db	2086	AGTTTAACTACAGGTTTGGCCCTCTTTTGTATACGAGCTGGAAAAATATATAGACAA	2145
OY	516	GAATTAAGATATTTTGGACAACGTTGCAACAGATTTGCACATTTGTTATGCGCAAAAAAAGAT	575
Db	2146	ATATAGAAAGTATTTCAATATTTTACGACAACTTCTCATTTGTCAAGTCTCAGGCTGAAAT	2205
OY	576	AGGATCTGGGTTTGATGTTCGCAACTGCACATTTTATGTCGTGATTTATATAGAAATTTCA	635
Db	2206	TGGAAACCGGTTTGATGTGAGCGCGCGAGCAATATGATCTATACGATATAGAAAGTTCC	2265
OY	636	GCCAGCTTTGATTAATGACGTGTTCAAGTCTTAAGAAATGATCCTGAGAAAGTCCCCAC	695
Db	2266	AACCGCATTAATCTTAATTTTCCAGATTT-----GGAAGTGTACTTATCGGCACTPA	2319
OY	696	AGAGTTGAAAAAATGATTTGAAGAAAGTAACGTGGGAATTTCAACATGAAGATGTACATTAC	755
Db	2320	ACTGCGCGATTTGTTGTGATGAAGAAAGACTGGAATATATACGATTAAGAAATTAACATTTAC	2379
OY	756	ATAGGAAATCAAGTTATTAATGCGGTGACGTCAAGGTGCGTCAAAAACCCCAATTTGCT	815
Db	2380	TTCCGGATTAATCTTTATGATAGCGGAGATTTTAAGATAGTTCACAGAAACGTPAAACGTGT	2439
OY	816	ATCAAGAGTACTCCAAATGAAAAAAGAAAGAACCCAGAAAGAAAGCTCTGTGTGTATGACCA	875
Db	2440	CCAGAAAGGTAAAAAATGTGTATGATTTGCAATATGCGCAAGAAAGCTTGAAAAATATATACGA	2499
OY	876	GCTTAATAGTGCATTTTACAGTTTATAGAAAGATT	911
Db	2500	ACTGATATATGCAAAATTTCTAGATTTTATGATGAGACT	2535

```

RESULT 11
US-10-006-909-12
; Sequence 12, Application US/100065909
; Publication No. US20030148479A1
; GENERAL INFORMATION:
; APPLICANT: KEASLING, JAY
; APPLICANT: MARTIN, VINCENT
; APPLICANT: PITERA, DOUGLAS
; APPLICANT: KIM, SEON-WON
; APPLICANT: WITHERS II, SYDNOR T.
; APPLICANT: YOSHIKUNI, YASUO
; APPLICANT: NEWMAN, JACK
; APPLICANT: KHEBNIKOV, ATEM VALENTINOVICH
; TITLE OF INVENTION: BIOSYNTHESIS OF ISOIDENTENTYL PYROPHOSPHATE
; FILE REFERENCE: 2000-0007
; CURRENT APPLICATION NUMBER: US/10/006,909
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 12
; LENGTH: 5051
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

```

```

; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: "M1" operon nucleotide sequence
US-10-006-909-12

```

Query Match	10.2%	Score 132;	DB 15;	Length 5051;
Best Local Similarity	50.5%;	Pred. No. 8.4e-23;		
Matches 473; Conservative	0;	Mismatches 430;	Indels 33;	Gaps 5

Qy	5	OTCAAAAGCATTTTATGTGCACTCGGAAAAAGCATTTCTTGCTGCTGTGATATTGGTCTTGA	62
Db	1606	GTGAGAGCCCTTCAGTGGCCCGAGGAAAGCCCTTCTAGCTGTGTGATATTGATTTTGA	1665
Qy	63	GCAATTTATATATCTTATGTGACAGCATTTGTCAACGAATGCAATGCAAGTTATTAACACC	122
Db	1666	TACAAATATGACATTTTGTATGTGCGATTTATCGGCAAGAAATGCAATGCTGAGCCCATCC	1722
Qy	123	AAAAGAAC-----CAGTTGAAAAGATCTGAAATCAAAATTTCTTACCCCAATTGC	176
Db	1726	TTTACGGTTTCATGTGCAAGGGCTGTGATTAAGTTGAAGTGGGTGTGAAAAGTAAACAATTTAA	1785
Qy	177	AAACCGAGAAATGGGAATATCAATATCATCAATAATACG--AGAACCCAGAGAAATTCA	233
Db	1786	AGATGGGAGAGTGGCTGATACATATATAGCTTAAAGTGGCTTCATCTCTGTTTCGATAGG	1845
Qy	234	GTCAACGATTAATCATTTTATGAGGCAACATATATATGCTTTAGCTTATTTATCAAC	293
Db	1846	CGGATCTTAAAGACCTTTTCATTTGAAAAAGTTATGCTTAACTATTTAGCTTATTAACC	1905
Qy	294	GACCGAAGCATTTGATCTTGAA-----TCATCATTTACTCAAGCCCTGGAATATCA	344
Db	1906	TAAATGAGACGACTACTGCAATAGAAATCTTGTTGTTATTTGATTTTCTCTGATGATGC	1965
Qy	345	TTTCAACAAGAAATCTGAAACCAAGACATCTCGAATGAGAAAAAATTTCTTTTACCA	404
Db	1966	CTACCATCTTCAGAGAGATAGCGTTATCCGAACATCTGGCAACGAAGATTTGAATTTTCA	2025
Qy	405	TTCTCGTCCATTAACGAAAGTGAAGAAACCGGATTAAGTTTCATCTGGCAGGATTAAGTTC	464
Db	2026	TTTGCACAGAAATTTGAAGAAAGTTCCCAAAACAGGCTGGCTCTCTGGAGGTTTATGTCAC	2085
Qy	465	AGTTGTTCCACAAGTTTATTTATCCATTTTATCCC-----CAATGTTATCAGTAC	515
Db	2086	AGTTTTAACATACAGCTTTGGCGTCTTTTGTGATGGAACCTGGAAAAATATATGAGCAA	2145
Qy	516	GAATAAAGATATTTTGCACAACGTTGCACAGATTTGCATTTGTATGCCAAAAAAAGAT	575
Db	2146	ATATAGGAAGATTATTCATATTTATTTAGACAAAGTTGCTCATTTGTCAACCTCAGGGTAAAT	2205
Qy	576	AGGATCTGGGTTGATTTGCAACTGCAATTTATGTCTGATTTGATATGAAATTTCA	635
Db	2206	TGGAAAGCGGTTTGTATGTAGCGCGGACGATATGATCTATTCAGATATAGAAATTTCC	2265
Qy	636	GCCAGCTTTGATTAATGACGTGTTCAGTTCAGAAAGTATCCTGAGAAAGTTCCCAAC	695
Db	2266	ACCGCATTTATCTCTATTTTGGCAAGATTT-----GGAAGTCTACTTAACGCAATGA	2319
Qy	696	AGAGTTGAAAAAATGATTTGAAAGTAACTGGGAATTAACAATGAAGATGATCAATTAAC	755
Db	2320	ACTGGCCCATTTGGTTGATGAAGAAAGCTGGAATATATACATTTAAAGTAAACCATTTTAC	2379
Qy	756	ATACGGAATCAAGTTATTAATGSGTGCACGCAAGGGGCGTCAGAAACCCCAATTTGGT	815
Db	2380	TTTGGGATTTAATTTATGATGAGGCGATTTTAAAGATGGTTTCAGAAAACATTAACATGGT	2439
Qy	816	ATCACAGTATCTCCAATTTGAAAAAGAAAGCAAGAAAGAGCTCTTGTGTATGACCA	875
Db	2440	CCGAAGGTAAAAAATTTGATATGCTTCGCAATATGCCAGAAAGCTTGAATAATATATACGA	2499
Qy	876	GCTTATATAGTCCCAATTTACAGTTTATGAAAGAAATT	911
Db	2500	ACTCGATCATGCAAAATTTGATTTATGATGATGACCT	2535

RESULT 12
US-10-411-066-12
; Sequence 12, Application US/10411066
; Publication No. US20040005678A1
; GENERAL INFORMATION:
; APPLICANT: KEASLING, JAY
; APPLICANT: MARTIN, VINCENT
; APPLICANT: PITERA, DOUGLAS
; APPLICANT: WITHERS III, SYDNOR T.
; APPLICANT: NEWMAN, JACK
; TITLE OF INVENTION: BIOSYNTHESIS OF AMORPHA-4,11-DIENE
; FILE REFERENCE: 2000-0007.20
; CURRENT APPLICATION NUMBER: US/10/411, 066
; PRIOR FILING DATE: 2000-04-09
; PRIOR APPLICATION NUMBER: 10/006, 909
; PRIOR FILING DATE: 2001-12-06
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 5051
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: "MBI" operon nucleotide sequence
US-10-411-066-12

Query Match 10.2%; Score 132; DB 16; Length 5051;
Best Local Similarity 50.5%; Pred. No. 8.4e-23;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCMAAAGCATTTAGTGCACCTGGAAAAGCATTTCTGTGTGTGATATTGGTCTTGA 62
DB 1606 GTTGAAGGCTTCAAGTCCCGGAGAAAGGCTTACGTGCGATATTAGTTTGA 1665
QY 63 GCCAATTATGATGCTTATGTGACAGCATTTGCATCGAATGCAATGCACTTAAACCC 122
DB 1666 TACAAATATGAAGCATTTGTAGTCGATATATCGGCAAGATCATGCTGTACCCATCC 1725
QY 123 AAAAGAAC-----CAGTTGAAAGATCTAGATCAAAATTTCTTACCCCAATTTGC 176
DB 1726 TTACGGTTATGCAAGGCTGTATGATTTGAAGTGGTGTGAAAGATTAACAAATTTAA 1785
QY 177 AAACGAGATGGGAATATCATATCATCAATAATACAG--AGAAAGCCAGAGAATTC 233
DB 1786 AGATGGGAGTGGCTGTACATTAAGTCTTAAAGGCTTCACTTCTTTCGATAGG 1845
QY 234 GTCACGCATTAATTCATTTTGAAGGCACATATATTCATGTTTGAAGTTTCAACC 293
DB 1846 CGGATCTAAGAACCTTTTCATGAAAAGTTATCGTAAAGTATTAGTAACTTAAAC 1905
QY 294 GACCGAAGCATTTGATCTTGAA-----TCATCATTTACTCAGACCTCGATATCA 344
DB 1906 TAACATGACGACTACTGTGCAATAGAAACTTGTTCGTTATGATATTTCTCTGATGATGC 1965
QY 345 TTCACAGAAGATGACTGAAACCAAGACATCTCGAATGAGAAAACATTTCTTAAACA 404
DB 1966 CTACCATTTCTCAGAGAGATAGCTTACCAACATCTGTGCAACAGAGATTGATTTCA 2025
QY 405 TTCTCGTCATTAACGAAGTGAAGAAACCGGATTAAGTTTCATCGGACAGATTAGTGC 464
DB 2026 TTTCGACAGAAATGAAGAGTTCCCAAAACAGGGCTGGGCTCTCGGACAGTTTAACTAC 2085
QY 465 AGTTGTGCGACAAAGTTTATATCCATTTTATCC-----CAATGTTATCAAGTAC 515
DB 2086 AGTTTAACTACAGCTTGTGCTCTTTTGTATGAGACCTGGAATAATATGTAAGCA 2145
QY 516 GAATTAAGATATTTGACAAAGTGGACAGATTCAGATTTGATGCCCCAAAAAGAT 575
DB 2146 ATTAAGAGATTTATTAATTTAGCAAGATTTGCTCATTTGTCAGAGCTTAAAGGTAAAT 2205
QY 576 AGGATCTGGTTGATGTTGCAACTGCAATTTATGCTGATTTGATATAGAAATTTCA 635
|||||

DB 2206 TGGAGCGGGTTTATGATGACGGCGGACAGATATGATCTATCATATTAAGAAATTTCC 2265
QY 636 GCCAGCTTGAATTAATAGAGTGTTCAGGTTCTAGAAAGTATCTGAGAAAGTTCCCA 695
DB 2266 ACCCGCATTAATCTCTAATTTTGCAGATATTT-----GGAAGCTACTTACCGCAGTAA 2319
QY 696 AGAGTTGAAAAAATTGATTGAAAGTAACTGGGAATTCAAACATGAAAGATTAATTAC 755
DB 2320 ACTGGCGCATTTGGTTGATGAGAAAGACTGGAATTAAGATTAAGAAAGTAAACATTTAC 2379
QY 756 ATACGAATCAAGTTTATATGAGTACGTCAGAGGTTGCTCGAAGAACCCAAATTTGT 815
DB 2380 TTGGGATTAATCTTATGATGAGTGGCGATTTAAGAAATGTTTCGAAACAGTAAACCTGT 2439
QY 816 ATCAGAGTACTCAATGAGAAAAAGAAAGCCAGAAAGAAAGCTGTGTGTATGACCA 875
DB 2440 CCAGAAAGTTAAAAAATTTGATATTCGATATGCGATATGCCAGAAAGCTTGAATAATATACAGA 2499
QY 876 GCTTAATAGTCCCAATTTACAGTTTATGAGAAAT 911
DB 2500 ACTGATCATGCAAAATTTCAATTTATGATGACT 2535
|||||

RESULT 13
US-10-006-909-13
; Sequence 13, Application US/10006909
; Publication No. US20030148479A1
; GENERAL INFORMATION:
; APPLICANT: KEASLING, JAY
; APPLICANT: MARTIN, VINCENT
; APPLICANT: PITERA, DOUGLAS
; APPLICANT: KIM, SEON-MON
; APPLICANT: WITHERS III, SYDNOR T.
; APPLICANT: YOSHIKUNI, YASUO
; APPLICANT: NEWMAN, JACK
; TITLE OF INVENTION: KHEBINIKOV, ARTEM VALENTINOVICH
; TITLE OF INVENTION: BIOSYNTHESIS OF ISOPENENTYL PYROPHOSPHATE
; FILE REFERENCE: 2000-0007
; CURRENT APPLICATION NUMBER: US/10/006, 909
; CURRENT FILING DATE: 2002-04-02
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 5963
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: "MBIS" operon nucleotide sequence
US-10-006-909-13

Query Match 10.2%; Score 132; DB 15; Length 5963;
Best Local Similarity 50.5%; Pred. No. 9.1e-23;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCMAAAGCATTTAGTGCACCTGGAAAAGCATTTCTGTGTGTGATATTGGTCTTGA 62
DB 1606 GTTGAAGGCTTCAAGTCCCGGAGAAAGGCTTACGTGCGATATTAGTTTGA 1665
QY 63 GCCAATTATGATGCTTATGTGACAGCATTTGCATCGAATGCAATGCACTTAAACCC 122
DB 1666 TACAAATATGAAGCATTTGTAGTGGATTTATGCGAAGATGCAATGCTGTACCCATCC 1725
QY 123 AAAAGAAC-----CAGTTGAAAGATCTAGATCAAAATTTCTTACCCCAATTTGC 176
DB 1726 TTACGGTTATGCAAGGCTGTATGATTTGAAGTGGTGTGAAAGATTAACAAATTTAA 1785
QY 177 AAACGAGATGGGAATATCATATCATCAATAATACAG--AGAAAGCCAGAGAATTC 233
DB 1786 AGATGGGAGTGGCTGTACATTAAGTCTTAAAGTGGCTTCACTTCTTTCGATAGG 1845
QY 234 GTCACGCATTAATTCATTTTGAAGGCACATATATTCATGTTTGAAGTTTCAACC 293
|||||

Db	1846	CGAGTCTAAGAACCCCTTTCATTGAAAAAGTTATCGTAAAGTATTTAGCTACTTTAAAC	1905
Qy	294	GACCGAAGCAATTGATCTTGAAA-----TCATCATTTTCTGAGACCTTGAAATATCA	344
Db	1906	TAAACATGGAAGCACTACTGCAATAGAAACTTGTCGTTATTTGATTAATTTTCTCGATGATGC	1965
Qy	345	TTACACAGAAGATATCTGAAACCAAGCATCTTCGATGAGAGAAAAACATTTCTTTACCA	404
Db	1966	CTACCAATTTCTCAGAGAGATAGCGTTACCGAACAATCGTGGCAACAGAAATTTGAGTTTCA	2025
Qy	405	TTCTCGTCCATTACCGAAGTGGAAAAAGACCGGATTTAGTTGATGGGAGAGTTAGTCTC	464
Db	2026	TTGCGACAGAAATTGAAGAAAGTTCCAAAACAGGGCTGGGCTCTCGCAGAGTTTAGTCAC	2085
Qy	465	AGTTGTTCACAAAGTTTATATATCCATTTTATCCC-----CAATGTTATCAGTAC	515
Db	2086	AGTTTATACCTACAGCTTTTGAGCCCTCTTTTGTATTCGCACTGMAAAATTAATGTAGCA	2145
Qy	516	GATAAAGATATTTTGGCAACAAGTGGACAGATTGSCATGTTATGGCCAAAAAAGAT	575
Db	2146	ATATAGAGAGATTATTCATATTTAGACAAAGTTGCTCATTTGTCAAGCTCAGGGTAAAT	2205
Qy	576	AGGATCTGGGTTTGATGTTGCAACTGCAATTATATGTCGATTTGATATAGAGATTTC	635
Db	2206	TGMAACGGGTTTGATGTAGCGCGGACGACATATGATCTATCAGATATAGAAATTTCC	2265
Qy	636	GCCAGCTTGATTAATATGACGTGTTCAAGTTCTAGAAAGTATCTTGAAAGTTCCCAAC	695
Db	2266	ACCCGATTAATCTCTAATTTGCGAGATATTT-----GGAAGTCTACTTACCGCAGTAA	2319
Qy	696	AGAGTTGAAAAAATGATTGAAAGTAACTGGGAATTCMAACATGAAAGATGATCATTACC	755
Db	2320	ACTGGCGCATTTGTTGTAAGAAAGAACTGGAATATTAAGATTAAAGTAAACATTTTACC	2379
Qy	756	ATACGGATCAAGTTTAAATGGGTGACGTCAAGGGTGGCTCAGAAAACACCAAAATTTGT	815
Db	2380	TTGGGATTAATCTTTATGATGAGGCGGATTTAAAGATGGTTCAAGAAACGTAAACTGG	2439
Qy	816	ATCACGAGTACTCCATGGAAGAAAGAAAGCAAGAAAGAAAGCTGTGTGTATGACCA	875
Db	2440	CCAGAAGTTAAAAAATTGATGATGATGCGCATATGCCGAAAAAGCTTGAAATATATTCAGA	2499
Qy	876	GCTTAATAGTGCCAATTTCAGTTTATGAAGAAATTT	911
Db	2500	ACTCGATCATCAAAATTCAGATTATGATGATGACT	2535
RESULT 14			
US-10-411-066-13			
Sequence 13, Application US/10411066			
Publication No. US20040005678A1			
GENERAL INFORMATION:			
APPLICANT: KEASLING, JAY			
APPLICANT: MARTIN, VINCENT			
APPLICANT: PITERA, DOUGLAS			
APPLICANT: WITHERS III, SYDOR T.			
APPLICANT: NEWMAN, JACK			
TITLE OF INVENTION: BIOSYNTHESIS OF AMORPHA-4,11-DIENE			
FILE REFERENCE: 2000-000720			
CURRENT APPLICATION NUMBER: US/10/411,066			
PRIOR FILING DATE: 2000-04-09			
PRIOR APPLICATION NUMBER: 10/006,909			
NUMBER OF SEQ ID NOS: 124			
SOFTWARE: PatentIn Ver. 2.1			
SEQ ID NO 13			
LENGTH: 5963			
TYPE: DNA			
ORGANISM: Artificial Sequence			
FEATURE:			
OTHER INFORMATION: Description of Artificial Sequence: Synthetic			
US-10-411-066-13			

Query Match	10.2%	Score 132;	DB 16;	Length 5963;
Best Local Similarity	50.5%	Pred. No. 9,1e-23;		
Matches	473;	Conservative	0;	Mismatch 430; Indels 33; Gaps 5;
Qy	3	GTCAAAAGCATTTAGTGAACCTGGAAAACATTTCTGTGGTGGATATTTGGTCTTGA	62	
Db	1606	GTTGAGAGCCTTCAGTCCCGCAGGAGAAACGTTACTAGCTGGTGGATATTTAGTTTGA	1666	
Qy	63	GCCAATTTAAGAGCTTATGTGACAGCATTTGTCTATCAGATATGACATGACATTAACAC	122	
Db	1666	TACAAATATGAAAGCATTTGTAGTCCGATTTATCCGCAAGAAAGCATGCTGTAGCCATCC	1725	
Qy	123	AAAAGAAC-----CAGTTGAAAGATCTAGATCAAAATTTCTCACCCCAATTTGC	176	
Db	1726	TTACGGTTATTCGCAAGGCTCTGTATAGTTTGAATGGCTGTGAAAGATTAACATTTTA	1785	
Qy	177	AAAGGAGAAATGGGAATATCAATATCATCAATACAG--AAGAGCCAGAGAAATTC	233	
Db	1786	AGATGGGAGGTGGCTGTACCATATATAGTCTTAAAGTGGCTTCATCTGTTCGATAGG	1845	
Qy	234	GTCAACCATTAATCCATTTTGTAGAGCAACTATATTCATCGTTTATAGCTTATATCAAC	293	
Db	1846	CGGATCTAAGAACCTTTTCATTTGAAAAGTTATGCTTAACTATTTAGCTTATTAAC	1905	
Qy	294	GACCGAAGCATTTGATCTTGAAA-----TCATCTTTTACTCAAGCCCTGGATATCA	344	
Db	1906	TAACATGAGCAGTCTGCAATATGAAACCTTGCTGTTATGATATTTTCTGTATGATGC	1965	
Qy	345	TTCAACAGAAAGATPACTGAAACCAAGACATCTCGAATGGAGAAAACATTTCTTTACA	404	
Db	1966	CTACACTTCTCAGAGAGATAGCGTTACCGAACATCTGTGGCAACGAAGATTAAGATTTC	2025	
Qy	405	TTCTGTGTCATTTACCGAAGTGGAAAGACCGGATTAAGTTTCACTGGCAGAGATTAAGTGC	464	
Db	2026	TTTCGACAGAAATGAAAGATTTCCCAAAACAGGCTGGGCTCGCGCAGGTTATGTAC	2085	
Qy	465	AGTTTGTCCCAAGTTTATATTCATCCATTTTATCCC-----CAATGTTATCAGTAC	515	
Db	2086	AGTTTATCTACAGTTTGGCTCTCTTTTGTATCGAACCTGGAAATATATGTAGCA	2145	
Qy	516	GAATTAAGATATTTTGCACAACTGTGCAGATTTGATGATGATGATGATGATGATGATG	575	
Db	2146	ATATGAAAGATTTATTCATATTTAGACAAAGTTGCTCATTTGCAAGCTCAGGGTAAAT	2205	
Qy	576	AGATCTGGGTTGATGTGTGCAACATGAAATTTATGCTATGTATATATGAAAGATTTC	635	
Db	2206	TGGAAAGGGGTTTATGTATGAGCGCGGACATATGATCTATAGATATATGAAAGATTTC	2265	
Qy	636	GCCAGCTTTGATTAATGACGTTTCAAGTTCTAGAAAGTATCTCGAAGATTTCCAC	695	
Db	2266	ACCGGCAATATCTCTAATTTTGCAGATATTT-----GGAAGTGCTAATTAAGGCAATTA	2319	
Qy	696	AGAGTTGAAAAAATGATGAAAGTAACTGGGAATTCAAACATGAAAGATATGATTAAC	755	
Db	2320	ACTGCGCATTTGGTTATGATGAAAGCTGGAATATTTAGCATTTAAAGTAACCATTTAC	2379	
Qy	756	ATACGGAATCAAGTTATTAATGAGTGAAGTCAAGGCTGCTCAGAAACCAACCAATTTGT	815	
Db	2380	TTGGGATTAACCTTATGATGAGGCGATATTAAGATGTTTCAAGAAACAGTAAACTGAT	2439	
Qy	816	ATCAGAGTACTCCAAATGGAAGAAAGAAAGCCAGAAAGAGCTCTGTTGTATGACCA	875	
Db	2440	CCAGAGATTAAGAAATTTGGTATGATTTCCGATATGCGCATATGCCAGAAAGCTGAAATATATACAG	2499	
Qy	876	GCTTAATAGTCCATTTTACAGTTTATGAAAGATTT	911	
Db	2500	ACTGATCATGCAAAATTTCTAGATTTATGATGATGACT	2535	

```
Publication No. US20030033626A1
GENERAL INFORMATION:
APPLICANT: Hahn, Frederick
APPLICANT: Kuehnle, Adelheid
TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
FILE REFERENCE: KAS-103XC1
CURRENT APPLICATION NUMBER: US/09/918,740
CURRENT FILING DATE: 2001-07-31
PRIORITY FILING DATE: 2000-07-31
PRIORITY FILING DATE: 2000-07-31
NUMBER OF SEQ ID NOS: 76
SOFTWARE: Patencin version 3.0
SEQ ID NO 61
LENGTH: 7681
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Operon C containing A. thaliana, S. cerevisiae, and Streptomyces
OTHER INFORMATION: CL190
OTHER INFORMATION: DNA, and R. capsulatus DNA
US-09-918-740-61
```

```
Query Match      10.2%; Score 132; DB 10; Length 7681;
Best Local Similarity 50.5%; Pred. No. 1e-22;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;
```

```
3 GTCAAAAGCATTTAGTGCACCTGGAAGAAAGCATTTCTGCTGGTGAATTTGGTTCTTGA 62
43 GTTGGAGGCTTCAGAGCCCGGAGAAACGTTACTAGCTGGTGAATTTAGTTTAA 102
63 GCCAATTTATGATGCTTATGTGACAGCATTTGTCATGACGATGATGATTAACACC 122
103 TACAAATATGAAAGCTTTGTAGTCGATTTATGCGAATGCAAGATGCTGTAGCCATCC 162
123 AAAAGGAC-----CAGTTGAAAGATCTAGAAATCTTCAACCCCAATTTGC 176
163 TTACGCTTATGCAAGGCTGATTAAGTTGAAGCGGTGAAAGTAAACAAATTTAA 222
177 AAACGGAATGGGAATATCACTATCATCAATTAACAG--AGAACCCAGAGAATGCA 233
223 AGATGGGAGTGGCTGTACCATTAATAGTCTTAAAGTGGCTTCATTCCTGTTGATAG 282
234 GTACGCAATTAATCCATTTTATAGGCACTATATTCATCGTTTATGCTTAATTCAC 293
283 CGGATTAAGAACCTTTCAATGAAAAGTTATCGCTAACGATTTAGCTACTTTAAAC 342
294 GACCGAAGCATTTGATCTTGAAA-----TCATCATTTTACTCAGACCCTGATATCA 344
343 TAAACATGACGACTACTGCAATAGAAATGTTTCGTTATGATTAATTTCTGATGATGC 402
345 TTCACAAAGAACTAGTAAACCAAGACATCTCGAATGAGAAAACAAATTTCTTACCA 404
403 CTACCATTTCTCAGAGAGATAGCGTTTACCAACATCGTGGCAACAGAAATGAGTTTCA 462
405 TTCTGTCGCAATTAACGAGTGGAAAAGCCGATTAAGTTTCTCGGACGATTAAGTGC 464
463 TTCGCAAGAAATGAAAGTTCCCAAAACAGGGCTGGGCTCTCGGACGTTTATGATC 522
465 AGTTGTCGCAAGTTTATTAATCCATTTTATCCC-----CAATGTTATCAGTAC 515
523 AGTTTAACTACAGCTTTGGCCCTCTTTTGTATGAGACCTGGAATAATATAGACAA 582
516 GAATTAAGATATTTTGCACACGTTGCAAGATTTGCACTGTTTATGCCCCAAAAAGAT 575
583 ATATAGAGAAGTATTCATTAATTTAGACAAAGTTGCTCATTTCAAGCTCAGGGTAAAT 642
576 AGGATCTGGTTGATGTTGCACTGCAATTTATGCTGATGTTATATAGATTTCA 635
643 TGGAGCGGGTTGATGATGAGCGGCGAGACATATGATCTATCAAGATTAAGAAATCCC 702
636 GCGAGCTTGAATTAATGACGTGTTCAAGTTCTAGAAAGTATCTGAGAAATTTCCAC 695
```

```
Db 703 ACCCGCATTAATCTCTAATTTCCAGATTT-----GGAAGTCTACTTACGCGATGA 756
Qy 696 AGAGTTGAAAAAATGATTAAGAAATGACGGAATTCACAAACATGAAGATGATACATTACC 755
Db 757 ACTGGCGCATTTGGTTGATGAAGAAAGCTGGAATTTAGATTAAGAAACCATTTAC 816
Qy 756 ATACGAATCAAGTTATTAATGAGTGAAGCTCAAGGCTGCTAGAAACCCAAATTTGGT 815
Db 817 TTCGGGATTAACCTTATGATGAGGCGATTAAGAAAGTGGTTCAGAAACAGTAAACTGGT 876
Qy 816 ATCAGAGTATCTCCAAATGAAAAAGAAAAGCAAGAAAGCTGTTGTATGACCA 875
Db 877 CCAGAGATTAAGAAATTTGATGATTCGATATGCCAGAAACCTGAAATATATACAG 936
Qy 876 GCTTAATAGTCCAAATTTACGTTATGAGGAAT 911
Db 937 ACTGCATCATGCAAAATTCATGATTTATGATGATGACT 972
```

RESULT 16

```
US-10-835-516-61
Sequence 61, Application US/10835516
Publication No. US20040194162A1
```

```
GENERAL INFORMATION:
APPLICANT: Hahn, Frederick
APPLICANT: Kuehnle, Adelheid
TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
FILE REFERENCE: KAS-103XC1
CURRENT APPLICATION NUMBER: US/10/835,516
CURRENT FILING DATE: 2004-04-28
PRIORITY FILING DATE: US/09/918,740
PRIORITY FILING DATE: 2001-07-31
PRIORITY FILING DATE: 2000-07-31
PRIORITY FILING DATE: 2000-07-31
NUMBER OF SEQ ID NOS: 76
SOFTWARE: Patencin version 3.0
SEQ ID NO 61
LENGTH: 7681
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Operon C containing A. thaliana, S. cerevisiae, and Streptomyces
OTHER INFORMATION: CL190
OTHER INFORMATION: DNA, and R. capsulatus DNA
US-10-835-516-61
```

```
Query Match      10.2%; Score 132; DB 18; Length 7681;
Best Local Similarity 50.5%; Pred. No. 1e-22;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;
```

```
3 GTCAAAAGCATTTAGTGCACCTGGAAGAAAGCATTTCTGCTGGTGAATTTGGTTCTTGA 62
43 GTTGGAGGCTTCAGAGCCCGGAGAAACGTTACTAGCTGGTGAATTTAGTTTAA 102
63 GCCAATTTATGATGCTTATGTGACAGCATTTGTCATGACGATGATGATTAACACC 122
103 TACAAATATGAAAGCTTTGTAGTCGATTTATGCGAATGCAAGATGCTGTAGCCATCC 162
123 AAAAGGAC-----CAGTTGAAAGATCTAGAAATCTTCAACCCCAATTTGC 176
163 TTACGCTTATGCAAGGCTGATTAAGTTGAAGCGGTGAAAGTAAACAAATTTAA 222
177 AAACGGAATGGGAATATCACTATCATCAATTAACAG--AGAACCCAGAGAATGCA 233
223 AGATGGGAGTGGCTGTACCATTAATAGTCTTAAAGTGGCTTCATTCCTGTTGATAG 282
234 GTACGCAATTAATCCATTTTATAGGCACTATATTCATCGTTTATGCTTAATTCAC 293
283 CGGATTAAGAACCTTTCAATGAAAAGTTATCGCTAACGATTTAGCTACTTTAAAC 342
294 GACCGAAGCATTTGATCTTGAAA-----TCATCATTTTACTCAGACCCTGATATCA 344
```



```
/ CURRENT APPLICATION NUMBER: US/10/835,516
/ CURRENT FILING DATE: 2004-04-28
/ PRIOR APPLICATION NUMBER: US/09/918,740
/ PRIOR FILING DATE: 2001-07-31
/ PRIOR APPLICATION NUMBER: 60/221,703
/ PRIOR FILING DATE: 2000-07-31
/ NUMBER OF SEQ ID NOS: 76
/ SOFTWARE: Patentin version 3.0
/ SEQ ID NO 58
/ LENGTH: 7693
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Operon containing A. thaliana and S. cerevisiae DNA
US-10-835-516-58

Query Match      10.2%; Score 132; DB 18; Length 7693;
Best Local Similarity 50.5%; Pred. No. 1e-22;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCAAAAGCATTTAGTGCACCTGGAAAAGCATTTCTTGCTGTGTGATTTGGTCTTGA 62
DB 3796 GTTGAAGCCTTCAGTGGCCCGAGGAAAGCGTTACTAGCTGGTGAATTTAGTTTGA 3855
QY 63 GCCAATTTATGATGCTTATGTGACAGCATTTGCATGACGAATGCAGTTTAAACACC 122
DB 3856 TACAAATATGAAGCATTTGTAGTGGATTTATCGGAGAAATGCATGCTGACCCATCC 3915
QY 123 AAAAGAAC-----CAGTTTGAAGAAATCTAGATCAAAATTTCTTCAACCCCAATTGC 176
DB 3916 TTAAGGTTATGCAAGGAGTGTGATTAAGTTGAAGCGGTGTAAGAAATTTAA 3975
QY 177 AAACGGAATGGGATATCATCATCATCAATATACG---AGAGCCGAGAGAATGTA 233
DB 3976 AGATGGAGTGGCTGTACCATTAATAGTCTTAAAGTGGCTTATCTGTTTCAATAG 4035
QY 234 GTACGCAATTAATTCATTTTATAGGCAACTATATTCATGTTTATGCTTATATTCAC 293
DB 4036 CGGATCTAAGAACCTTTTCAATGAAAAGTTATCGTAAAGTTATAGCTTAAAC 4095
QY 294 GACCGAAGCATTTGATCTTGA---TCATCATTTACTCAGACCTTGATATCA 344
DB 4096 TAAACATGACGACTACTGCAATGAAACTTGTGCTGATTTGATTTTCTCATGATGC 4155
QY 345 TTCACAGAAGATCTAGAAACCAAGCATCTCGAATGAGAAAATTTCTTTACCA 404
DB 4156 CTACCATTTCTCAGAGATAGCGTTACCGAATCGTGGCAAGAAATTTGAGTTTCA 4215
QY 405 TTCTCGTGCATTTACGGAAGTGAAGAACCGGATTAAGTTTCATCGGACGATTAAGTGC 464
DB 4216 TTGCAACAGATTTGAAGAGTTCCCAAAACAGGGCTGGCTCTCGGACAGTTTATGTCAC 4275
QY 465 AGTTGTTGCCACAGTTTATTTATCCATTTATCCC-----CAATGTTATCAGTAC 515
DB 4276 AGTTTATTAACAGCTTTGGCCCTCTTTTGTATGAGCACTGGAATAATATGATAGCA 4335
QY 516 GAATTAAGATTTTTCAGACAGCTTGCACAGATTTGCACATTTTATGCCCAAAAAGAT 575
DB 4336 ATATTAAGAAATTTATCATTAATTTAGCAAGTTGCTCATTTGCAAGCTCAGGGTAAAT 4395
QY 576 AGGATCTGGGTTTATGTTGCACTGCAATTTATGCTGATTTATATGAGATTTCA 635
DB 4396 TGGAAACGGGTTTATGATGAGCGGCGGACATATGATCTATCAGATTAAGAAATTTCC 4455
QY 636 GCCAGCTTTGATTAATGACGCTTTTCAAGTGAAGATGATCTGAGAAAGTTCCCCAC 695
DB 4456 ACCCGATTAATCTCTAATTTGCCAGATAT-----GGAAGTCTACTTACCGCAGTAA 4509
QY 696 AAGTTGAAGAAATTTGATGAAGTAACTGGGAATTTCAACATGAAGATGATCATTAAC 755
DB 4510 ACTGGCGCATTTGTTGATGAGAAAGATGGAATTTATTCAGATTAAGATCAATTAAC 4569
QY 756 ATACGAATCAAGTTTATTAATGAGTGAAGTCAAGGGTGGCTCAGAAACCCAAATTTG 815
```

```
DB 4570 TTGCGGATTAATCTTATGATGGCGCATATTTAAGATGGTTGCAAAACGTAAACCTG 4629
QY 816 ATACAGATTAATCTTATGATGAGAAAGCAAGAAAGCAAGAAAGCTGTTGATAGCA 875
DB 4630 CCAGAAAGTAAATAATTTGATGATTCGCAATATGCAAGAAAGCTTGAATATATACGA 4689
QY 876 GCTTAATAGTGCATTTTACAGTTTATGAGAAATTT 911
DB 4690 ACTGCATATGCAAAATTTTATGATGAGTGA 4725

RESULT 19
US-09-918-740-59
/ Sequence 59, Application US/09918740
/ Publication No. US20030033626A1
/ GENERAL INFORMATION:
/ APPLICANT: Hahn, Frederick
/ TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
/ FILE REFERENCE: KAS-103XCI
/ CURRENT APPLICATION NUMBER: US/09/918,740
/ PRIOR FILING DATE: 2001-07-31
/ PRIOR APPLICATION NUMBER: 60/221,703
/ NUMBER OF SEQ ID NOS: 76
/ SOFTWARE: Patentin version 3.0
/ SEQ ID NO 59
/ LENGTH: 7695
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Operon B containing A. thaliana and S. cerevisiae DNA
US-09-918-740-59

Query Match      10.2%; Score 132; DB 10; Length 7695;
Best Local Similarity 50.5%; Pred. No. 1e-22;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCAAAAGCATTTAGTGCACCTGGAAAAGCATTTCTTGCTGTGTGATTTGGTCTTGA 62
DB 28 GTTGAAGCCTTCAGTGGCCCGAGGAAAGCGTTACTAGCTGGTGAATTTAGTTTGA 87
QY 63 GCCAATTTATGATGCTTATGTGACAGCATTTGCATGACGAATGCAGTTTAAACACC 122
DB 88 TACAAATATGAAGCATTTGTAGTGGATTTATCGGAAATGCAATGCTGATGCCATTC 147
QY 123 AAAAGAAC-----CAGTTTGAAGAAATCTAGATCAAAATTTCTTCAACCCCAATTGC 176
DB 148 TTAAGGTTATGCAAGGCTGTGATTAAGTTGAAGTGGCTGTGAAGAAATTAACATTTAA 207
QY 177 AAACGGAATGGGAAATATCATCATCATCAATATACG---AGAGCCGAGAGAATGTA 233
DB 208 AGATGGGAGTGGCTGTACCATTAATAGTCTTAAAGTGGCTTCATTCGTTTGCATAG 267
QY 234 GTACGCAATTAATTCATTTTATAGGCAACTATATTCATGCTTTATAGTTTATTCAC 293
DB 268 CGGATCTAAGAACCTTTTATTAAGAAAGTTATCGTAAACGTAATTTAGCTTAAAC 327
QY 294 GACCGAAGCATTTGATCTTGAA---TCATCATTTACTCAGACCTTGATATCA 344
DB 328 TAAACATGACGACTACTGCAATTAAGAACTTTGCTGTTATTTATGATTTATGATGATGC 387
QY 345 TTCACAGAAGATTAAGTGAAGCAAGCATCTCGAATGAGAAATAATTTCTTTTACCA 404
DB 388 CTACCATTTCTCAGAGATAGCGTTTACCAACATCGTGGCAAGAAAGTTGAGTTTCA 447
QY 405 TTCTCGTGCATTTACGGAAGTGAAGAACCGGATTAAGTTATGCTGATGATGATGC 464
DB 448 TTGCAACAGATTTGAAGAGTTCCCAAAACAGGGCTGGCTCTCGGACAGTTTATGTCAC 507
QY 465 AGTTGTTGCCACAGTTTATTTATCCATTTTATCCC-----CAATGTTATCAGTAC 515
```

```

Db      508 AGTTTAACTACAGCTTGGCCCTCTTTTGTATCGAAGCTGGAATATATAGACAA 567
QY      516 GAATAAAGATATTTTGCACACAGTTGACAGATTTGCACATTTGATATGCCAAAAAAGAT 575
Db      568 ATATAGAGAAATTTCTAATTTTAAAGCAAGATGCTCATTTGCTAACCTCAGGGTAAAT 627
QY      576 AGAATCTGGTTTGTATGTTGCACTGCAATTTATGATCTGATTTATATAGAAATTCA 635
Db      628 TGAAGCGGGTTTATGATGAGCGGCGGAGATATGAGATCTATCGAATATAGAAATTTCC 687
QY      636 GCCAGCTTTGATTAATGACGTTTCAAGTTCTAGAAAGTATCTGAGAAAGTTCCAC 695
Db      688 ACCCGATTAATCTCTAATTTGCGAGATATT-----GGAAGTCTCTCTTACGCGAGTAA 741
QY      696 AGAGTTGAAAAAATTTGATTTGAAAGTAACTGGGAAATTCAACATGAAGATGATACATTAC 755
Db      742 ACTGGCCCATTTGGTTATGAGAAAGCTGGAATATTACGATTTAAAGTAAACCATTTACC 801
QY      756 ATACGGAATCAAGTTATTTAAATGGGTGACGTCAAGGGTGGCTCAGAAAACCCAAATTTGT 815
Db      802 TTCCGGGATTTAATCTTATGATGAGCGCGATTTAAGAAATGTTTCAGAAACAGTAAACTGGT 861
QY      816 ATACAGATCTCCAAATGGAAGAAAGCAAGAAAGCTCTGTTGTATGACCA 875
Db      862 CCGAAGGTAAAAAATTTGATGATTCGATATGCCAGAAAGCTTGAAATATATACAGA 921
QY      876 GCTTAATAGTCCAAATTTACGTTTATGAGAAATT 911
Db      922 ACTGATCATGCAAAATTTCTAGATTTATGATGAGACT 957

RESULT 20
US-10-835-516-59
; Sequence 59, Application US/10835516
; Publication No. US20040194162A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Frederick
; APPLICANT: Kuehne, Adelheid
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
; FILE REFERENCE: KAS-103XCI
; CURRENT APPLICATION NUMBER: US/10/835,516
; CURRENT FILING DATE: 2004-04-28
; PRIOR APPLICATION NUMBER: US/09/918,740
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/221,703
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 59
; LENGTH: 7695
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Opecon B containing A. thaliana and S. cerevisiae DNA
US-10-835-516-59

Query Match      10.2%; Score 132; DB 18; Length 7695;
Best Local Similarity 50.5%; Pred. No. 1e-22;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

```

```

QY      177 AAACGGAATGGGAATATCATATCATCAATACAG---AGAAACCCAGAAAGTTCA 233
Db      208 AGATGGGAGTGGCTGTATCATATATATATATATATATATATATATATATATATATATATAT 267
QY      234 GTTACGCAATTAATTCATTTTATAGAGCAATATATATATATATATATATATATATATATATAT 293
Db      268 CGATCTAAGAAACCTTTCATTTGAAAGATATATATATATATATATATATATATATATATATAT 327
QY      294 GACCGAAGCATTTGATCTTGA---TCATCATTTTATCTGAGACCTGATATCA 344
Db      328 TAAATGACGACATCTGCAATAGAAAGCTGTTGCTATATATATATATATATATATATATATAT 387
QY      345 TTCACAGAAGATATCTGAACCAACATATCTCCGAAATGAGAAAGAAATATATCTTTACCA 404
Db      388 CTACATTTCTCAGAGATAGCTTTACCAATCTGCAACATCTGCAACAGAAATGAGATTTTCA 447
QY      405 TTCTGTGTCATTAACCAAGTGAAGAACGAGATTTAGTTTATGATGCGAGATTTATGTC 464
Db      448 TTGCGACAGAAATTTGAAGAAAGTTCCAAAACAGGCTGCGGCTCTGCGAGGTTATGTCAC 507
QY      465 AGTTGTGCGCAAGTTTATATATCCATTTTATCC-----CAATGTATCAGTAC 515
Db      508 AGTTTAACTACAGCTTTGGCCCTCTTTTGTATCGAAGCTGGAATATATAGACAA 567
QY      516 GAATAAAGATATTTTGCACACGTTGACAGATTTGCAATTTGATATGCCAAAAAAGAT 575
Db      568 ATATAGAGAAATTTATTTCAATTTTAAAGTCAAGTGTCTATTTGCAAGCTCAGGGTAAAT 627
QY      576 AGAATCTGGTTTGTATGTTGCACTGCAATTTATGATGATTTATATAGAAATTCA 635
Db      628 TGAAGCGGGTTTGTATGATGAGCGGCGGAGATATGATCTATGATATATAGAAATTTCC 687
QY      636 GCCAGCTTTGATTAATGACGTTTCAAGTTCTAGAAAGTATCTGAGAAAGTTCCAC 695
Db      688 ACCCGATTAATCTCTAATTTGCGAGATATT-----GGAAGTCTCTCTTACGCGAGTAA 741
QY      696 AGAGTTGAAAAAATTTGATTTGAAAGTAACTGGGAAATTCAACATGAAGATGATACATTAC 755
Db      742 ACTGGCCCATTTGGTTATGAGAAAGCTGGAATATTACGATTTAAAGTAAACCATTTACC 801
QY      756 ATACGGAATCAAGTTATTTAAATGGGTGACGTCAAGGGTGGCTCAGAAAACCCAAATTTGT 815
Db      802 TTCCGGGATTTAATCTTATGATGAGCGCGATTTAAGAAATGTTTCAGAAACAGTAAACTGGT 861
QY      816 ATACAGATCTCCAAATGGAAGAAAGCAAGAAAGCTCTGTTGTATGACCA 875
Db      862 CCGAAGGTAAAAAATTTGATGATTCGATATGCCAGAAAGCTTGAAATATATACAGA 921
QY      876 GCTTAATAGTCCAAATTTACGTTTATGAGAAATT 911
Db      922 ACTGATCATGCAAAATTTCTAGATTTATGATGAGACT 957

RESULT 21
US-09-918-740-62
; Sequence 62, Application US/09918740
; Publication No. US2003003626A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Frederick
; APPLICANT: Kuehne, Adelheid
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
; FILE REFERENCE: KAS-103XCI
; CURRENT APPLICATION NUMBER: US/09/918,740
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/221,703
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 62
; LENGTH: 8224
; TYPE: DNA

```

ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Operon E containing A. thaliana, S. cerevisiae, Steptomyces sp C
 OTHER INFORMATION: DNA,
 OTHER INFORMATION: and R. capsulatus
 US-09-918-740-62

Query Match 10.2%; Score 132; DB 10; Length 8224;
 Best Local Similarity 50.5%; Pred. No. 1.1e-22;
 Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCAAAAGCATTTAGTGCACCTGGAAGCATTTCTGCTGGTGGATTTGGTTCTTGA 62
 DB 43 GTTGAGAGCCTTCAGTCCAGGGAAGGCTTACGCTGGGATATTAGTTTGA 102
 QY 63 GCCAATTTATGATGCTATGTGA CAGCATGTCAATCGAATGATGCACTTTAAACCC 122
 DB 103 TACAAATATGAAAGCATTTTGTAGTGGATTTAGCGAATGATGCTTACCCATCC 162
 QY 123 AAAAGGAAAC-----CAGTTGAAAGATCTAGATCAAAATTTCTTCAACCCCATTTGC 176
 DB 163 TTACGGTTATTCAGAGGCTGATTAAGTTGAAGTCGCTGAAAGTAAACATTTAA 222
 QY 177 AAAAGGAAATGGGAATATCAATATCAATCAATACAG--AGAAAGCCAGAGAGTTCA 223
 DB 223 AGATGGGAGTGGCTGATACATTAAGTCTTAAAGGCTTCAATTCCTGTTTCGATAGG 282
 QY 234 GTACACCATTAATTCATTTTATGAGGCACTATATTCATGCTTTAGCTTAAATTCAC 293
 DB 283 CGGATCTAAGAACCTTTCAATTAAAGTTATGCTTAACGATTTAGCTTTAAAC 342
 QY 294 GACCGAAGCATTTGATCTTGAAA-----TCATCATTTACTCAGACCTGGATATCA 344
 DB 343 TAAACATGAGAGACTACGCAATAGAACTGTGTTGATTTGATTTTCTCTAGATGTC 402
 QY 345 TTCAACAAGAAATCTGAAACCAAGACATCTCGAATGAGAAAGAAACATTTCTTAA 404
 DB 403 CTACCATTTCTCAGAGGATAGCGTTACGAAACATCGGCAACAGAAATTTGAGTTTCA 462
 QY 405 TTCTCGTGCATTAACGGAATGGAAGAAAGCCGATTAAGTTGATCGGACAGATTAAGTC 464
 DB 463 TTGCAACAGATTAAGAAAGTTCCCAAAAGGCTGGCTCTCCGCAAGTTAGTCAC 522
 QY 465 AGTTGTTGCCACAAGTTTATATCCATTTATCC-----CAATGTTACAGTAC 515
 DB 523 AGTTTAACTACAGCTTTGGCTCTCTTTTGTATGCGACTGGAATTAATAGACAA 582
 QY 516 GAATTAAGATATTTTGACAACGTTGACAGATTTGACATTTGTTATGCCCAAAAAAGAT 575
 DB 583 ATATAGAGAAATTAATCATTAATTAGACAAGTTGCTCATTTCAAGCTCAGGGTAAAT 642
 QY 576 AGATATCGGTTGATGTTGCACTGCAATTTATGCTGATGTTATNTAGAAATTTCA 635
 DB 643 TGAAGCGGGTTGATGTTAGCGGCGGACAGATATGATCTATCAGATTAAGAAATTTCC 702
 QY 636 GCCAGCTTTGATTAATGACGTTTCAAGTTCTAGAAAGTATCCTGAGAAGTTCCAC 695
 DB 703 ACCCGCATTAATCTTAATTTGCCAGATTT-----GGAAGTCTACTTAACGAGTAA 756
 QY 696 AGAGTTGAAAAAATTTGATTAAGTAATCTGGAATTCMAACATGAAGATGTACATTACC 755
 DB 757 ACTGGGCAATTTGTTGATGTAAGAAAGCTGAAATATTAAGATTAACATTTTACC 816
 QY 756 ATACGAATCAATTTATTAATGCTGACGTCAGAGGCTGCAAGAAACACCAATTTGT 815
 DB 817 TTGGGATTTAACTTTATGATGGCGCATTAATTAAGAAAGTTTCAGAAACATTAACGCT 876
 QY 816 ATCAGAGTACTCAATGAAAAAGAAAGCAAGAAAGCTCTGTGTATGACCA 875
 DB 877 CCAGAAAGTAAAAATTTGATGATTTGCAATGCGCATATGCCAAAGCTTGAATATATACGA 936
 QY 876 GCTTAATAGTCCAAATTTTACAGTTTATGAAGAATT 911

DB 937 ACTCGATCATGCAAAATTTAGATTTATGATGATGACT 972

RESULT 22

US-10-835-516-62
 Sequence 62, Application US/10835516
 Publication No. US20040194162A1
 GENERAL INFORMATION:
 APPLICANT: Hahn, Frederick
 APPLICANT: Kuehle, Adelheid
 TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
 TITLE OF INVENTION: create novel traits in transgenic organisms
 FILE REFERENCE: KAS-103XCI
 CURRENT APPLICATION NUMBER: US/10/835,516
 PRIOR APPLICATION NUMBER: 2004-04-28
 PRIOR FILING DATE: 2001-07-31
 PRIOR APPLICATION NUMBER: 60/221,703
 PRIOR FILING DATE: 2000-07-31
 NUMBER OF SEQ ID NOS: 76
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 62
 LENGTH: 8224
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Operon E containing A. thaliana, S. cerevisiae, Steptomyces sp C
 OTHER INFORMATION: DNA,
 OTHER INFORMATION: and R. capsulatus
 US-10-835-516-62

Query Match 10.2%; Score 132; DB 18; Length 8224;
 Best Local Similarity 50.5%; Pred. No. 1.1e-22;
 Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCAAAAGCATTTAGTGCACCTGGAAGCATTTCTGCTGGTGGATTTGGTTCTTGA 62
 DB 43 GTTGAGAGCCTTCAGTCCAGGGAAGGCTTACGCTGGGATATTAGTTTGA 102
 QY 63 GCCAATTTATGATGCTATGTGA CAGCATGTCAATCGAATGATGCACTTTAAACCC 122
 DB 103 TACAAATATGAAAGCATTTTGTAGTGGATTTAGCGAATGATGCTTACCCATCC 162
 QY 123 AAAAGGAAAC-----CAGTTGAAAGATCTAGATCAAAATTTCTTCAACCCCATTTGC 176
 DB 163 TTACGGTTATTCAGAGGCTGATTAAGTTGAAGTCGCTGAAAGTAAACATTTAA 222
 QY 177 AAAAGGAAATGGGAATATCAATATCAATCAATACAG--AGAAAGCCAGAGAGTTCA 233
 DB 223 AGATGGGAGTGGCTGATACATTAAGTCTTAAAGGCTTCAATTCCTGTTTCGATAGG 282
 QY 234 GTACACCATTAATTCATTTTATGAGGCACTATATTCATGCTTTAGCTTAAATTCAC 293
 DB 283 CGGATCTAAGAACCTTTCAATTAAAGTTATGCTTAACGATTTAGCTTTAAAC 342
 QY 294 GACCGAAGCATTTGATCTTGAAA-----TCATCATTTACTCAGACCTGGATATCA 344
 DB 343 TAAACATGAGAGACTACGCAATAGAACTGTGTTGATTTGATTTTCTCTAGATGTC 402
 QY 345 TTCAACAAGAAATCTGAAACCAAGACATCTCGAATGAGAAAGAAACATTTCTTAA 404
 DB 403 CTACCATTTCTCAGAGGATAGCGTTACGAAACATCGGCAACAGAAATTTGAGTTTCA 462
 QY 405 TTCTCGTGCATTAACGGAATGGAAGAAAGCCGATTAAGTTGATCGGACAGATTAAGTC 464
 DB 463 TTGCAACAGATTAAGAAAGTTCCCAAAAGGCTGGCTCTCCGCAAGTTAGTCAC 522
 QY 465 AGTTGTTGCCACAAGTTTATATCCATTTATCC-----CAATGTTACAGTAC 515
 DB 523 AGTTTAACTACAGCTTTGGCTCTCTTTTGTATGCGACTGGAATTAATAGACAA 582
 QY 516 GAATTAAGATATTTTGACAACGTTGACAGATTTGACATTTGTTATGCCCAAAAAAGAT 575

Db 583 ATATAGAGAAATTATTCATATTTAGACAAAGTTGCTCATTTGCAAGCTCAGGGTAAAT 642
QY 576 AGGATCTGGGTTGATGTTGCAACTGCAATTTATGCTGATTTGATATAGAAATTTCA 635
Db 643 TGGAGCGGGTTGATGATGACGGGAGATATGATCTATGATATATGAAGATTTCC 702
QY 636 GCCAGCTTTGATTAATGACGTGTTTCAAGTTCTAGAAAAGTATCTGAGAGATTCCCC 695
Db 703 ACCCGATTAAATCTCTAATTTGGCAGATATTT-----GGAAAGTCTACTTACCGCAGTAA 756
QY 696 AGAGTTGAAAAAATGATGTAACCTGGAATTCAAACATGAAGATGATACATTTAC 755
Db 757 ACTGGGCGCATTTGGTGTATGATAGAAAGCTGAAATTTAGCATTTAAAGTAAACATTTAC 816
QY 756 ATACGGAATCAAGTTATTTAATGAGTGAAGTCAAGGGTGGCTCAGAAAACCCAAATTTGT 815
Db 817 TTGGGATTAATCTTTATGATGAGGCGATATTAAGAAATGTTTCAAGAAACGATAAACCTGT 876
QY 816 ATACAGATCTCCATGTAAGAAAAAGAAAAAGCAAGAAAGCTCTGTTGTATGACCA 875
Db 877 CCAGAAAGTAAAAAATTTGATATTTGCAATATGCCAAGAAAGCTGAAAAATATATACAGA 936
QY 876 GCTTAATAGTCCAAATTTACGTTTATGAGGAATT 911
Db 937 ACTGATCATGCAAAATTTCTAGATTTATGATGGAAT 972

RESULT 23

US-09-918-740-60
; Sequence 60, Application US/09918740
; Publication No. US20030033626A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Frederick
; APPLICANT: Kuehnle, Adelheid
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
; FILE REFERENCE: KAS-103XC1
; CURRENT APPLICATION NUMBER: US/09/918, 740
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/221, 703
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 60
; LENGTH: 8235
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Operon C containing A. thaliana, S. cerevisiae, and R. caosulatus
US-09-918-740-60

Query Match 10.2%; Score 132; DB 10; Length 8235;
Best Local Similarity 50.5%; Pred. No. 1,1e-22;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCAAAAGCATTTAGTGCACCTGGAAGCATTTCTGCTGTGATATTTGGTTCTGA 62
Db 28 GTTGAAGCGCTTCAAGTCCCGAGGGAAGCGTTACTAGCTGGGATATTTAGTTTAA 87
QY 63 GCCAATTTATGATGCTATGTCACAGCATTTGCATACGAAGGCAAGCTATTAACAC 122
Db 88 TACAAATATGAAAGCATTTGTAGTGGATTTATCGGAAGATGCAATGCTAGCCATCC 147
QY 123 AAAAGAAC-----CAGTTTGAAGAAATCTAGATCAAAATTTCTTCAACCCCAATTTGC 176
Db 148 TTAAGGTTTATGTCAGAGGCTGTATAGTTTGAAGCGGTGTAAGAAAGTAAACAAATTTAA 207
QY 177 AAAAGGAATGGAATATTCATATCAATCAATCAATAGAG---AGAACCCAGAGAAATTTCA 233
Db 208 AGATGGGAGTGGCTGTACATATTAAGTCTTAAAGTGCTTCAATCTGTTTCAATAG 267
QY 234 GTACAGCATTAATTCATTTTATAGAGCACTATATTCATGCTTTATATTTCAACC 293

Db 268 CCGATCTAAGAACCTTTTCATTGAAAAAGTATTCGTAACGATATTTAGCTATTAAACC 327
QY 294 GACCGAAGCATTTGATCTTGAA-----TCATCATTTTACTGACACCTGTGATATCA 344
Db 328 TAAATAGACGACTACTGCAATAGAAACCTTTGCTTATGATATTTTCTGATGATGC 387
QY 345 TTCAAGAAATATCTGAAACCAAGACATCTCGAATGAGAAAAAATTTCTTTACCA 404
Db 388 CTACCATCTTCAGAGAGATAGGTTTACCAACATCGTGGCAACAGAAATGATTTTCA 447
QY 405 TTCTGTCGATTAACCGAATGGAAGAACCGGATTAAGTTGATGGCGAGATTTGTC 464
Db 448 TTGCAAGAAATGAAAGTTCACAAACAGGGCTGGCTCTCGGCGGTTTGTGTCAC 507
QY 465 AGTTGTCACAAAGTTTATTTATCCATTTTATCC-----CAATGTTATCAATAC 515
Db 508 AGTTTATCAATGACTTTGGCTCTCTTTTGTATGACCTGGAAGATATGATAGACA 567
QY 516 GAATAAAGTATTTTGCACACGTTGCACAGATGCAATTTGATATGCCCCAAAAAGAT 575
Db 568 ATATAGAGAAATTTATTCATTAATTTAGCAAGATGCTCATGTGCAAGCTCAGGGTAAAT 627
QY 576 AGGATCTGGGTTGATGTTGCAACTGCAATTTATGCTGATTTATATAGAAATTTCA 635
Db 628 TGAAGCGGGTTGATGATGACGGCGGCGAGCATATGATCTATCAATATAGAAATTTCC 687
QY 636 GCCAGCTTGAATTAATGACGTGTTCAAGTTCTAGAAAGTGAATCTGAGAAATTTCC 695
Db 688 ACCCGCATTAATCTCTAATTTGCCAGATATTT-----GGAAGTCTACTTACGGCAGTAA 741
QY 696 AGAGTTGAAAAAATTTGATGAAAGTGAATCTGGAATTTCAACATGAAAGATTTAC 755
Db 742 ACTGGCGCATTTGGTGTATGAAAGAACTGGAATTTTCAATTAAGTAAACCAATTTAC 801
QY 756 ATACGGAATCAAGTATTAATGAGTGAAGTGAAGGTTGCTCAGAAACCCAAATTTGT 815
Db 802 TTGGGATTAATCTTTATGATGAGGCGATATTAAGATGTTTCAAGAAACGTAATCTGT 861
QY 816 ATACAGATCTCCATGTAAGAAAAAGAAAAAGCAAGAAAGCTCTGTTGTATGACCA 875
Db 862 CCAGAAAGTAAAAAATTTGATATGATTCGATATGCAAGAAAGCTTGAATATATACAGA 921
QY 876 GCTTAATAGTCCAAATTTTACGTTTATGAGGAATT 911
Db 922 ACTGATCATGCAAAATTTCTAGATTTATGATGGAAT 957

RESULT 24

US-10-835-516-60
; Sequence 60, Application US/10835516
; Publication No. US20040194162A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Frederick
; APPLICANT: Kuehnle, Adelheid
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
; FILE REFERENCE: KAS-103XC1
; CURRENT APPLICATION NUMBER: US/10/835, 516
; PRIOR FILING DATE: 2004-04-28
; PRIOR APPLICATION NUMBER: US/09/918, 740
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/221, 703
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 60
; LENGTH: 8235
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Operon C containing A. thaliana, S. cerevisiae, and R. caosulatus
US-10-835-516-60

Query Match 10.2%; Score 132; DB 18; Length 8235;
 Best Local Similarity 50.5%; Pred. No. 1.1e-22;
 Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCAAAAGCATTTAGTGCACCTTGAAAAGCATTTCTTGCTGGTGGATTTGGTTCTTGA 62
 DB 28 GTTGAGAGCCTTCAGGCCCGCAGGAAAGGCTTACTAGCTGGGATTTAGTTTGA 87
 QY 63 GCCAATTTATGATGCTATGTGACAGATGTCATGCAAGATGCAATGCAATGTTAAACACC 122
 DB 88 TACAAATATGAAAGCATTTGATGCTGATTAATGCGAAGATGATGCTTACCCATCC 147
 QY 123 AAAAGAAC-----CAGTTGAAAGATCTAGATCAAAATTTCTTACCCCATTTGC 176
 DB 148 TTACGGTTCTTCAGAGGGCTGATTAAGTTGAAGTCGGTGAAGAAAGTAAACATTTAA 207
 QY 177 AAACGAGATGGAAATATGATATCATCAATTAACAG--AGAACCCAGAGAGTTCA 233
 DB 208 AGATGGGAGTGGCTGATCAATTAAGTCTTAAAGTGGCTTCAATTCCTGTTTCGATAGG 267
 QY 234 GTCAAGCATTAATTCATTTTAAAGGCACTATATTCATGCTTTAGCTTATATTCAC 293
 DB 268 CGGATTTAAGAACCTTTTCACTGAAAAGTTATCGCTAACGATTTAGCTTAAAC 327
 QY 294 GACCGAAGCATTTGATCTTGAAA-----TCATCATTTACTCAGACCCCTGATATCA 344
 DB 328 TAAATGAGAGCACTACGCAATAGAAACTGTTGCTGATTTGATTTCTCTGATATGC 387
 QY 345 TTCAACAAGATGATCTGAAACCAAGACATCTCGAATGAGAAAACATTTCTTTACCA 404
 DB 388 CTACCATTTCTCAGAGGATAGCGTTACCGAACATCGGCGAACAGAAATGAGTTTCA 447
 QY 405 TTCTCGGCATTAACGAGAGTGGAAAAGACCGGATTAAGTTTCAATCGGCAAGATTAAGTC 464
 DB 448 TTGCAACAGATTAAGAAAGATTTCCCAAAACAGGCTGGCTCTCGCAGAGTTAGCAC 507
 QY 465 AGTTGTGCAACAAGTTTATATCCATTTTATCCC-----CAATGTTATCACTAC 515
 DB 508 AGTTTAACTACAGCTTTGGCTCTTTTGTATGAGACCTGAAAATATATGAGCA 567
 QY 516 GAATTAAGATTTTGGCAACAGTTGCAAGATTTGATGTTTATGCCCCAAAAAGAT 575
 DB 568 ATATAGAGAAATTAATCATTAATTAGCAACAAGTTGCTCATTTCAAGCTCAGGTAAT 627
 QY 576 AGGATGCTGTTGATGTTGCACTGCAATTTATGCTGATGTAATGAAATTTCA 635
 DB 628 TGAAGCGGTTTGAATGAGCGGCGCAGCATATGATCTATCAGATATGAAAGATTTCC 687
 QY 636 GCCAGCTTGAATTAAGACGTGTTCAAGTTCTAGAAAGTATCTGAGAAATTTCCAC 695
 DB 688 ACCCGATTAATCTTAATTTGCCAGATAT-----GGAAGTCTACTTAACGAGTAA 741
 QY 696 AGAGTTGAAAAATTAATGAAAGTAACTGGGAATTCAAACATGAAAGATGATCAATACC 755
 DB 742 ACTGGGCGCTTTGATGATGAAGACCTGGAATATTAAGATTAACCAATTTACC 801
 QY 756 ATACGAATCAATTTTAAATGAGTGAAGTCAAGGCTGCTCAAGAAACCCCAATTTGT 815
 DB 802 TTGGAATTAATCTTAATGAGTGAAGTGAATTAAGAAATGTTCAAGAAACAGTAAATCGT 861
 QY 816 ATCAGAGTACTCAATGGAAGAAAGAAAGCAAGAAAGCTCTGTTGTATGACCA 875
 DB 862 CCAAGAGTAAATTTGATGATGCTGATGCTATGCGCAAGAAAGCTTGAATATATACAG 921
 QY 876 GCTTAATAGTCCAAATTTACAGTTTATGAAGAAAT 911
 DB 922 ACTGATCATGCAAAATTTAGATTTATGATGAGTACT 957

RESULT 25
 US-09-918-740-64
 ; Sequence 64, Application US/09918740
 ; Publication No. US20030033626A1

GENERAL INFORMATION:
 APPLICANT: Hahn, Frederick
 APPLICANT: Kuehne, Adelheid
 TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
 TITLE OF INVENTION: create novel traits in transgenic organisms
 FILE REFERENCE: KAS-103X1
 CURRENT FILING DATE: 2001-07-31
 PRIOR APPLICATION NUMBER: 60/221,703
 PRIOR FILING DATE: 2000-07-31
 NUMBER OF SEQ ID NOS: 76
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 64
 LENGTH: 8400
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURe: Other Information: Operon G containing A. thaliana, S. cerevisiae, and S. pombe DNA
 US-09-918-740-64

Query Match 10.2%; Score 132; DB 10; Length 8400;
 Best Local Similarity 50.5%; Pred. No. 1.1e-22;
 Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

QY 3 GTCAAAAGCATTTAGTGCACCTTGAAAAGCATTTCTTGCTGGTGGATTTGGTTCTTGA 62
 DB 28 GTTGAGAGCCTTCAGGCCCGCAGGAAAGGCTTACTAGCTGGGATTTAGTTTGA 87
 QY 63 GCCAATTTATGATGCTATGTGACAGATGTCATGCAAGATGCAATGCAATGTTAAACACC 122
 DB 88 TACAAATATGAAAGCATTTGATGCTGATTAATGCGAAGATGATGCTTACCCATCC 147
 QY 123 AAAAGAAC-----CAGTTGAAAGATCTAGATCAAAATTTCTTACCCCATTTGC 176
 DB 148 TTACGGTTCTTCAGAGGGCTGATTAAGTTGAAGTCGGTGAAGAAAGTAAACATTTAA 207
 QY 177 AAACGAGATGGAAATATGATGCTGCAATTCATCAATTAACG--AGAACCCAGAGAGTTCA 233
 DB 208 AGATGGGAGTGGCTGATCAATTAAGTCTTAAAGTGGCTCATTCCTGTTTCGATAGG 267
 QY 234 GTCAAGCATTAATTCATTTTAAAGGCACTATATTCATGCTTTATGCTTATATTCAC 293
 DB 268 CGGATTTAAGAACCTTTTCACTGAAAAGTTATCGCTAACGATTTAGCTTAAAC 327
 QY 294 GACCGAAGCATTTGATCTTGAAA-----TCATCATTTACTCAGACCCCTGATATCA 344
 DB 328 TAAATGAGAGCACTACGCAATAGAAACTGTTGCTGATTTGATTTCTCTGATATGC 387
 QY 345 TTCAACAAGATGATCTGAAACCAAGACATCTCGAATGAGAAAACATTTCTTTACCA 404
 DB 388 CTACCATTTCTCAGAGGATAGCGTTACCGAACATCGTGGCAACAGAAATTTGAGTTTCA 447
 QY 405 TTCTCGGCATTAACGAGAGTGGAAAAGACCGGATTAAGTTTCAATCGGCAAGATTAAGTC 464
 DB 448 TTGCAACAGATTAAGAAAGATTTCCCAAAACAGGCTGGCTCTCGCAGAGTTAGCAC 507
 QY 465 AGTTGTGCAACAAGTTTATATCCATTTTATCCC-----CAATGTTATCACTAC 515
 DB 508 AGTTTAACTACAGCTTTGGCTCTTTTGTATGAGACCTGGAATATATGATGAGCA 567
 QY 516 GAATTAAGATTTTGGCAACAGTTGCAAGATTTGCAATTTGATGCTTAAACAAAAAGAT 575
 DB 568 ATATAGAGAAATTAATCATTAATTAGCAACAAGTTGCTCATTTGCAAGCTCAGGTAAT 627
 QY 576 AGGATGCTGTTGATGTTGCACTGCAATTTATGCTGATGTAATGAAAGATTTCA 635
 DB 628 TGAAGCGGTTTGAATGAGCGGCGCAGCATATGATCTATCAGATATGAAAGATTTCC 687
 QY 636 GCCAGCTTGAATTAAGACGTGTTCAAGTTCTAGAAAGTATCTGAGAAATTTCCAC 695
 DB 688 ACCCGATTAATCTTAATTTGCCAGATAT-----GGAAGTCTACTTAACGAGTAA 741
 QY 696 AGAGTTGAAAAATTAATGAAAGTAACTGGGAATTCAAACATGAAAGATGATCAATACC 755

```
Db      742 ACTGGCGCATTTGGTTGATGAAGAGACTGGAAATATTAGCAATTAAAGTAACCATTTTACC 801
Qy      756 ATACGGAATCAAGTTATTAATGAGGTGACGTCAAGAGGTGCTCAGAAACCCCAATTGGT 815
Db      802 TTCGGGATTAACCTTTATGATGGCGCATTTAAGAAATGGTTGAGAAACGATAAACCTGGT 861
Qy      816 ATCAGCAGTACTCCAAATGGAAGAAAGAACCCAGAGAAAGCTGTGTGTATGACCA 875
Db      862 CCAGAGATTAAGAAAAATTTGGTATGATTCGCATATGCCAGAAAGCTTGAATAATATACAGA 921
Qy      876 GCTTAATAGTCCCAATTTTACAGTTTATGAAGAATT 911
Db      922 ACTGCATCATGCAAAATTTAGATTATGATGAGACT 957

RESULT 26
US-10-835-516-64
; Sequence 64, Application US/10835516
; Publication No. US20040194162A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Frederick
; APPLICANT: Kuehnle, Adelheid
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
; FILE REFERENCE: KAS-103XC1
; CURRENT APPLICATION NUMBER: US/10/835,516
; CURRENT FILING DATE: 2004-04-28
; PRIOR APPLICATION NUMBER: US/09/918,740
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/221,703
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 64
; LENGTH: 8400
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Operon G containing A. thaliana, S. cerevisiae, and S. pombe DNA
US-10-835-516-64

Query Match      10.2%; Score 132; DB 18; Length 8400;
Best Local Similarity 50.5%; Pred. No. 1.1e-22;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

Qy      3  GTCAAAAGCATTTAGTGAACCTGGAAAGCATTTCTGTGTGATTTAGTTGTTCTTGA 62
Db      28  GTTGAGAGCTTCAGTGGCCCAAGGAAAGCGTTACTAGCTGGGATTTAGTTTGA 87
Qy      63  GCCAATTTATGATGCTTATGTGACAGCATTTGTCATCAGCAATGATGACAGTTTAAACCC 122
Db      88  TACAATAATGAAGCATTTGTAGTGGATTTATGCGAAGAAATGATGCTGTAGCCATGC 147
Qy      123  AAAAGAAC-----CAGTTTGAAGAATCTAGAAATCAAAATTTCTTCAACCCCAATTTGC 176
Db      148  TTACGGTTTATGCAAGGGCTGATTAAGTTTGAAGGCGGTGAAAGTAACAATTTAA 207
Qy      177  AAACGAGATGGGAATATCATCATCAATCAAG---AGAACCCAGAGAAATTTCA 233
Db      208  AAGTGGGAGTGGCTGTACATTAAGTCTTAAAGTGGCTTCATTCCTGTTCGATAGG 267
Qy      234  GTACAGCATTAATCCATTTTATAGGCAACTATATTATTCATGTTTATGCTTATTAACCC 293
Db      268  CGGATCTAAGAACCTTTTATGTAAGAAAGTATGCTAAGTATTTAGCTACTTTAAAC 327
Qy      294  GACCGAAGCATTTGATCTTGAAA-----TCATCATTACTCAGACCTGATATCA 344
Db      328  TAAACATGACGACTACTGCAATGAGAACTTGTTGTTATTTGATTTATTTCTGATGATGC 387
Qy      345  TTCACAGAAAGATCTGAAACCAAGCATCTCGAATGAGAGAAAAACATTTCTTTACA 404
Db      388  CTACCATTTCTCAGAGAGATAGCGTTTACGAAACATCGTGGCAACAGAAATGATGATTTCA 447
```

```
Qy      405  TTCGTGTCATTTACCGAAGTGGAAAAGACCGGATTTAGTTGATCGGACGATTAAGTGC 464
Db      448  TTCGACAGAAATTTGAAGAAATTTCCCAAAACAGGGCTGGGCTCTCGGCGAGTTTACTAC 507
Qy      465  AGTTGTGACACAAAGTTTATTTATCCCATTTTATCCC-----CAATGTTATCAGTAC 515
Db      508  AGTTTAACTTACAGCTTTGGCCCTCTTTTATTCGACCTGGAAATAATATGTAGACAA 567
Qy      516  GAATTAAGATATTTTGCACACAGTTGCACAGATTTGTCATTTATGCCCCAAAAAGAT 575
Db      568  ATATAGAGAGATTTATTTCAATTTAGCACAAATTTGCTCATTTGCAAGCTCAGGTAAT 627
Qy      576  AGAATCTGGGTTGATGTTGCAATGCAATTTATGCTGATTTGATATATAGAAATTTCA 635
Db      628  TGAAGCGGGTTGATGATGACGGCGGACGATATGATCTATCAGATATAGAAATTTCCC 687
Qy      636  GCCACCTTGAATTAATGACGTGTTACAGTTCTAGAAAGTATCCTGAGAAATTTCCCC 695
Db      688  ACCCGCATTAATCTCTAATTTGCCAGATATT-----GGAAGTCTACTTACGGCAGTAA 741
Qy      696  AGAGTTGAAAAAATTTGATGAAGTAACTGGGAATTTCAACATGAAGATGTACATTACC 755
Db      742  ACTGGCGCATTTGGTTGATGAGAAAGACTGGAATTTAGATTAAGTAACCATTTACC 801
Qy      756  ATACGGAATCAAGTTATTAATGGGTGACGTCAAGGGTGGCTCAGAAACCCCAATTGGT 815
Db      802  TTCGGGATTAACCTTTATGATGGCGCATTTAAGATGGTTTCAAGAAACGTAACCTGGT 861
Qy      816  ATCAGCAGTACTCCAAATGGAAGAAAGAACCCAGAGAAAGCTGTGTGTATGACCA 875
Db      862  CCAGAGATTAAGAAAAATTTGGTATGATTCGCTATGCGCAGAAACCTTGAAATAATATACAGA 921
Qy      876  GCTTAATAGTCCCAATTTTACAGTTTATGAAGAATT 911
Db      922  ACTGCATCATGCAAAATTTAGATTATGATGAGACT 957

RESULT 27
US-10-006-909-7
; Sequence 7, Application US/10006909
; Publication No. US20030148479A1
; GENERAL INFORMATION:
; APPLICANT: KEASLING, JAY
; APPLICANT: MARTIN, VINCENT
; APPLICANT: PITERA, DOUGLAS
; APPLICANT: WITHERS III, SYDNOR T.
; APPLICANT: YOSHIKUNI, YASUO
; APPLICANT: NEWMAN, JACK
; APPLICANT: KHEBENIKOV, ARTEM VALENTINOVICH
; TITLE OF INVENTION: BIOSYNTHESIS OF ISOBERENTYL PYROPHOSPHATE
; FILE REFERENCE: 2000-0007
; CURRENT APPLICATION NUMBER: US/10/006,909
; CURRENT FILING DATE: 2002-04-02
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 7
; LENGTH: 9253
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: "single operon" nucleotide sequence
US-10-006-909-7

Query Match      10.2%; Score 132; DB 15; Length 9253;
Best Local Similarity 50.5%; Pred. No. 1.2e-22;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

Qy      3  GTCAAAAGCATTTAGTGAACCTGGAAAGCATTTCTGTGTGATTTAGTTGTTCTTGA 62
Db      6259  GTTGAGAGCTTCAGTGGCCCAAGGAAAGCGTTACTAGCTGTGTGATTTATTTAGTTTGA 6318
```


Db 4557 TTCGACACGAAATTGAAGAAAGTTCCAAAACAGGCGCTGGGCTCCTCGAGGGTTAGTCAC 4616

Qy 465 AGTTGTTGCCACAAGTTTATTATCCATTTTATCC-----CAATGTTATCAGTAC 515

Db 4617 AGTTTAACTCAAGCTTGGGCTCTTTTGTATTCGAACTGGAAAAATATATGACAA 4676

Qy 516 GAATTAAGATATTTTGCACAACGTTGCACAGATTGCATTTGTATGCCAAAAAAGAT 575

Db 4677 ATATGAGAAATTATTCATTAATTAGCAACAAGTGTCTATTGTCAGCTCAGGGTAAAT 4736

Qy 576 AGAATCTGGGTTTGATGTTGCACATTCGAATTTATGCTCGATGTGTATTAAGATTCA 635

Db 4737 TGAAGACCGGGTTTGATGTAACGCGCGCACACATATGATCTATGAAATATGAAGATTCCC 4796

Qy 636 GCCAGCTTGAATTAATGACGTTTTCAGGGTCTAGAAAGTATCTGGAAGTTCCCCAC 695

Db 4797 ACCCGCATTAATCTCTAATTTTCCAGATATTT-----GGAAAGCTACTTACGGCAGTAA 4850

Qy 696 AGAGTTGAAAAAATGATGAAAGTAACCTGGGAATTCAAACATGAAAGATGTACATTACC 755

Db 4851 ACTGCGCATTTGTTGTTGATGAAGAAAGACTGGAATATTAACATTAATTAACCATTTACC 4910

Qy 756 ATACGGAATCAGTTATTTATGATGGTGAACGTCAAGGGTGTCTCAAGAACCCCAATTTGCT 815

Db 4911 TTGGGATTAACCTTATGATGATGGCGCATATTTAAGATGTTCTCAAGAACAGTAAACTGGT 4970

Qy 816 ATCAGAGTACTCCAATGAAAAAAGAAAAAGCAAGAAAGCTCTGTGTATGACCA 875

Db 4971 CCAGAAAGTAAAAAATTTGATATGATTCGATATGCCAAGAAAGCTTGAATAATATATACGA 5030

Qy 876 GCTTAATAGTCCCAATTTACAGTTTATGAAGAAAT 911

Db 5031 ACTGATCATGCAAAATTTAGATTATGATGAGACT 5066

RESULT 30

US-10-835-516-72

/ Sequence 72, Application US/10835516

/ Publication No. US20040194162A1

/ GENERAL INFORMATION:

/ APPLICANT: Hahn, Frederick

/ APPLICANT: Kuehler, Adelheid

/ TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways

/ TITLE OF INVENTION: create novel traits in transgenic organisms

/ FILE REFERENCE: KAS-103XC1

/ CURRENT APPLICATION NUMBER: US/10/835,516

/ PRIOR FILING DATE: 2004-04-28

/ PRIOR APPLICATION NUMBER: US/05/918,740

/ PRIOR FILING DATE: 2001-07-31

/ PRIOR APPLICATION NUMBER: 60/221,703

/ PRIOR FILING DATE: 2000-07-31

/ NUMBER OF SEQ ID NOS: 76

/ SOFTWARE: PatentIn version 3.0

/ SEQ ID NO 72

/ LENGTH: 13917

/ TYPE: DNA

/ ORGANISM: Artificial Sequence

/ FEATURE:

/ NAME/KEY: misc feature

/ LOCATION: (1)..(1)

/ OTHER INFORMATION: Plasmid transformation vector pKHO4, containing Operon B, containi

/ OTHER INFORMATION: 1

US-10-835-516-72

Query Match 10.2%; Score 132; DB 18; Length 13917;

Best Local Similarity 50.5%; Pred. No. 1.4e-22;

Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

Qy 3 GTCAAAAGCATTTAGTGCACCTGGAAGAAAGCATTTCTTGCTGTGATATTTGGTTCTTGA 62

Db 4137 GTTGAGAGCCTTCAGTGCACCCAGGAAAGCCTTACTAGCTGTGTGATATTTAGTTTAA 4196

Qy 63 GCCAATTATGATGCTTATATGACAGCATTTGTCATACGAATGACATGCAATATTAACACC 122

```

Db      4197 TACAAATATGAGCAATTTGTAGTCGGATTAATCGGCAAGATGATGCTAGCCATCC 4256
Qy      123 AAAAGAAC-----CAGTTGAAAGATCTAGAAATCTTACACCCCAATTGC 176
Db      4257 TTACGGTCTATGCAAGGCTCTGATTAAGTTGAGAGCCGTGAAAGATTAACAATTAA 4316
Qy      177 AACGGAAGATGGGAATATCATATCATCAATATACAG---AGAGCCAGAGAGTTCA 233
Db      4317 AGATGGGAGGTGCTGTACATATAAGTCTTAAAGGGCTTCACTCTGTTTCGATAGG 4376
Qy      234 GTACGCAATTAATCCATTTTAAAGGCACTATATATGTTTATGCTTATATTTCAAC 293
Db      4377 CGGATCTAAGAACCTTTCATTAAGAAAGTTATCGCTAAGATTTTGTACTTAAACC 4436
Qy      294 GACCGAAGCATTTGATCTTGA---TCATCATTTACAGACCCGTGATATCA 344
Db      4437 TAAATGAGAGCACTACGCAATAGAACTTGTTGATTAATATTTTCTGTATGATGC 4496
Qy      345 TTGACAGAAGATATCTGAAACCAAGATCTCGAATGAGAAACATTTCTTTACCA 404
Db      4497 CTACCAATTCACAGAGATAGCGTTACGAAACATCGTGCAACAGAAATTTGAGTTTCA 4556
Qy      405 TTCTCGTCCATTACCGAATGGAACCCGATTAGTTTCATCGGCAAGATTAGTGC 464
Db      4557 TTGCAACAGAAATGGAAGATCCCAAAACAGGCGTGGCTCCTCGGCAAGTTAGTAC 4616
Qy      465 AGTTGTTGCCAAGATTATATCCATTTTATCCC-----CAATGTTATCACTAC 515
Db      4617 AGTTTAACTACAGCTTTCGCTCTTTTGTATGCGACCTGAAATTAATATAGCA 4676
Qy      516 GAATAAAGATATTTTGCAACGTTGCAAGATTGCAATTTGATGCCCAAAAGAT 575
Db      4677 ATATAGAGAGATTATCATTAATTTAGCAAGATTGCTCATTTGCAAGCTCAGGTAAT 4736
Qy      576 AGGATCTGGTTGATGTTGCAACTGCAATTTATGCTGATTTGATATAGAAATTCA 635
Db      4737 TGGAAACGGGTTGATGATGACGGCGGCAAGATGATCATGATATGAAAGATTCCC 4796
Qy      636 GCCAGTTGATTAAGAGCTGTTTCAGTTCTAGAAAGATCCGAGAAAGTTCCCAAC 695
Db      4797 ACCCGATTATATCTTAATTTGCCAGATATTT-----GGAAGTGTACTTACGGCAGTAA 4850
Qy      696 AGAGTTGAAAAAATTTGATTAAGTAAGTAACGTGGAATTAACAATGAAGATGATACCTAC 755
Db      4851 ACTGGGCACTTGGTTGATTAAGTAAGTAAGTAATTTACATTAAGTAACATTTACC 4910
Qy      756 ATACGAATCAAGTTATTAATGAGTGAAGTCAAGGCTGCTCAGAAACACCCAAATTGCT 815
Db      4911 TTCGGGATTAACTTTATGATGAGGCGATATTAAAGATGTTCAAGAAACAGTAAACTGCT 4970
Qy      816 ATACAGAGTACTCAATGCAAAAAGAAAAGCAAGAAAGATCTCTGTGTATGACCA 875
Db      4971 CCAGAAAGTAAATAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 5030
Qy      876 GCTTAATAGTCAATTTACGTTTATGAAGAAATT 911
Db      5031 ACTGATCATGCAAAATTTCTAGATTTATGATGACT 5066

```

RESULT 31

```

US-09-918-740-74
; Sequence 74, Application US/0918740
; Publication No. US20030033626A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Frederick
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
; TITLE OF INVENTION: create novel traits in transgenic organisms
; FILE REFERENCE: KAS-103XCI
; CURRENT APPLICATION NUMBER: US/09/918,740
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/221,703

```

```

; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: Patent version 3.0
; SEQ ID NO 74
; LENGTH: 14623
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1..7)
; OTHER INFORMATION: Plasmid transformation vector pBK08, containing Operon G, contai
US-09-918-740-74

Query Match
Best local similarity 50.5%; Pred. No. 1,5e-22;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

Qy      3  GTCAAAAGCATTTAGTGCACCTGAAAGCAATTTCTGCTGTGATATTTGTTCTTGA 62
Db      2621 GTTGAAGCCTTCAGTGCCTCCAGGAAAGGCTTACTAGCTGTGATATTTAGTTTGA 2680
Qy      63  GCCAATTTATGATGCTTATGACAGATTGTGATACAGAAATGATGCACTTATACACC 122
Db      2681 TACAAATATGAAACAATTTGTAGTGGATTAATGCGAAGATCATGCTGTAGCCATCC 2740
Qy      123  AAAAGAAC-----CAGTTGAAAGATCTAGAAATCAAAATTTCTTACCCCAATTGC 176
Db      2741 TTACGGTCTATGCAAGGCTCTGATTAAGTTGAGAGCCGTGAAAGTAAACAATTAA 2800
Qy      177  AACGGAAGATGGGAATATCATATCATCAATATACAG---AGAGCCAGAGAAATTCA 233
Db      2801 AGATGGGAGGTGCTGTACATATAAGTCTTAAAGGCTTCATTCCTGTTTCATAGG 2860
Qy      234  GTCAAGCAATTAATCCATTTTAAAGGCACTATATTCATGTTTATGCTTATATTCACC 293
Db      2861 CGGATCTAAGAACCTTTCAATGAAAGATTAATCGCTAAGATTAATGCTTAAAC 2920
Qy      294  GACCGAAGCATTTGATCTTGA---TCATCATTTACTCAGACCCGTGATATCA 344
Db      2921 TAAATGAGAGCACTACGCAATTAAGAACTTGTTGTTATGATTAATTTCTCTGATGATGC 2980
Qy      345  TTGACAGAAGATATCTGAAACCAAGATCTCGAATGAGAAACATTTCTTTACCA 404
Db      2981 CTACCAATTCACAGAGATAGCTTACCAACATGCTGGAACAGAAAGATTTGATTTCA 3040
Qy      405  TTCTCGTCCATTACCGAAGTGAAGAACCCGATTAGTTCAATCGGCAAGATTAGTGC 464
Db      3041 TTGCAACAGAAATGAAAGATTTCCAAACAGGCGTGGCTCCTCGGCAAGTTAGTAC 3100
Qy      465  AGTTGTTGCCAAGTTTATATCCATTTTATCCC-----CAATGTTATCACTAC 515
Db      3101 AGTTTAACTACAGCTTTCGCTCTTTTGTATGAGACTGGAATTAATATAGCA 3160
Qy      516  GAATAAAGATATTTTGCAACGTTGCAAGATTGCAATTTGATGCTTAAAGAT 575
Db      3161 ATATAGAGAGATTATCATTAATTTAGCAAGATTTGCTCATTTGCAAGCTCAGGTAAT 3220
Qy      576  AGGATCTGGTTGATGTTGCAACTGCAATTTATGCTGATTTGATTAAGAAATTCA 635
Db      3221 TGGAAACGGGTTTGAATGAGCGGCGGCAAGCATATGATCATATATGAAAGATTCCC 3280
Qy      636  GCCAGTTGATTAAGAGCTGTTTACAGTTCTAAGAAAGATCCGAGAAAGTTCCCAAC 695
Db      3281 ACCCGATTATATCTTAATTTGCCAGATATTT-----GGAAGTGTACTTACGGCAGTAA 3334
Qy      696  AGAGTTGAAAAAATTTGATTAAGTAAGTAACGTGGAATTAACAATGAAGATGATACCTAC 755
Db      3335 ACTGGGCACTTGGTTGATTAAGTAAGTAAGTAATTTAGATTAAGTAACATTTACC 3394
Qy      756  ATACGAATCAAGTTATTAATGAGTGAAGTCAAGGCTGCTCAGAAACACCCAAATTGCT 815
Db      3395 TTGCGGATTAACTTTATGATGAGGCGATATTAAAGATGTTCAAGAAACAGTAAACTGCT 3454

```



```

Qy 123 AAAAGAAC-----CAGTTGAAAGATCTAGATCAAAATTTCTTCACCCCAATTGC 176
Db 2741 TTACGGTTCAATGCAAGGGCTGTGATTAAGTTGAAGTCGTGTGAAGTAACATTTAA 2800
Qy 177 AAAAGGAATGGGAATATCAATATCATCAATATACAG--AGAACCCAGAGATTCA 233
Db 2801 AGATGGGGAGTGGGTGACATATTAAGTCTTAAAGTGGCTTCAATCTGTTTCGATAGG 2860
Qy 234 GTACCGATTAATCCATTTTAAAGGCACTATATTCATGCTTTTACCTTAATTAACC 293
Db 2861 CGATCTTAAGAACCTTTCATGTAAGAAAGTTATGCTTAAGCTATTTAGCTAATTAAAC 2920
Qy 294 GACCGAAGCATTTGATCTTGA---TCATCATTTAATCTAGACCTCGATATCA 344
Db 2921 TTAATGAGAGACATACGCAATAGAAACTTGTCTTATTAATTTCTCTGTAGATGC 2980
Qy 345 TTCAACAAGATATCTGAACCAAGACATCTCGAATGAGAGAAAAATTTCTTTAACA 404
Db 2981 CTACCAATTTCTAGAGGATAGCGTTTACGAACATCTGCGCAACAGAGATTGAATTTTCA 3040
Qy 405 TTCTCGGCAATTAACCAAGTGAAGACCGGATTAAGTTTATCGGCAAGATTAGTGC 464
Db 3041 TTGCGACAGAATTGAAGAAAGTCCCAAAACAGGGCTGGCTCCCTCGGCAAGTTTACAC 3100
Qy 465 AGTTGTCGCAAGTTTATATATCCATTTTATCC-----CAATGTATCAGTAC 515
Db 3101 AGTTTAACTACGACTTTGGCTCTCTTTTGTATCGAACCCTGAAAAATATAGACAA 3160
Qy 516 GAAATAAAGATATTTTGACAAAGTTGACAGATTTGACATTTTATGCCCCAAAAAGAT 575
Db 3161 ATATAGAGAGATTAATCATTAATTTAGACAAAGTTGCTCATGTGCAAGCTCAGGGTAAAT 3220
Qy 576 AGAATCGGGTTTGAATGTTGCACTGCAATTTATGCTGATTTATATAGAAATTTCA 635
Db 3221 TGAAGCGGGTTGATGTAACGGCGGACGATATGATCTATCAAGATATGAAATTTCC 3280
Qy 636 GCCAGCTTTGATTAATGACGTTTCAAGTTCTTGAAGATGATCCTGAGAAAGTTCCAC 695
Db 3281 ACCCGATTAACTCTTAATTTGCGCAATTT-----GGAAGTCTTAATCGGCAATBA 3334
Qy 696 AGAATGAAAAAATTTGATTAAGATTAATCTGGAATTTCAACATGAAAGATGATCAATAC 755
Db 3335 ACTGGCGCATTTGGTTGATGAAGAAAGCTGGAATTTACGATTAAGTAACATTTACC 3394
Qy 756 ATACGGAATCAAGTTATTAATGAGTGAACGTCAGAGGGTCTCAGAAACCCCAATTTGT 815
Db 3395 TTGGGGATTAACTTTATGATGGCGATTTAAGAAATGTTTCAAGAAACAGTAAACTGGT 3454
Qy 816 ATACAGGATCTCAATGGAAGAAAGAAAGCCAGAGAAAGCTCTGTGTATGACCA 875
Db 3455 CCGAAGGTTAAAAAATTTGATGATTTGCGCATATGCCGAAAGCTTGAAGAAATATACGA 3514
Qy 876 GCTTAATAGTCCCAATTTTACAGTTTATGAAGAAATT 911
Db 3515 ACTGATCATGCAATTTCTAGATTTATGATGAGACT 3550

```

```

RESULT 34
US-10-835-516-76
; Sequence 76, Application US/10835516
; Publication No. US20040194162A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Frederick
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
; FILE REFERENCE: KXS-103XCI
; CURRENT APPLICATION NUMBER: US/10/835,516
; PRIOR FILING DATE: 2004-04-28
; PRIOR APPLICATION NUMBER: US/09/918,740
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/221,703

```

```

; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 76
; LENGTH: 14623
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1..7)
; OTHER INFORMATION: Plasmid transformation vector pFHO6, containing Operon E, contai
; OTHER INFORMATION: 1
US-10-835-516-76

Query Match      10.2%; Score 132; DB 18; Length 14623;
Best Local Similarity 50.5%; Pred. No. 1.5e-22;
Matches 473; Conservative 0; Mismatches 430; Indels 33; Gaps 5;

Qy 3 GTCAAAAGCATTTAGTGCACTGGAAGAAAGATTTCTGTGTGATATTTGTTCTTGA 62
Db 2621 GTTGAAGCCCTTCACTGCCCCAGAGAAAGGTTTACTAGCTGGTGAATATTTAGTTTGA 2680
Qy 63 GCCAATTTATGATGCTTATGTAACAGATTGTCAATCGAATGATCAAGTTATTAAC 122
Db 2681 TACAAATATGAAGCATTTGTATGTCGATTAATGCGCAAGATGATGATGATGCCATCC 2740
Qy 123 AAAAGAAC-----CAGTTGAAAGATCTAGATCAAAATTTCTCACCCCAATTGC 176
Db 2741 TTACGGTTCAATGCAAGGCTGATTAAGTTGAAGTCGTGTGAAGTAACAAATTTAA 2800
Qy 177 AAAAGGAATGGGAATATCAATATCATCAATATACAG--AGAACCCAGAGATTCA 233
Db 2801 AGATGGGAGTGGGTGACATATTAAGTCTTAAAGTGGCTCATTTCTGTTGATAGG 2860
Qy 234 GTACGCAATATTCATTTTAAAGGCACTATATTCATGTTTATGCTTAATTAACC 293
Db 2861 CGATCTTAAGAACCTTTCATGTAAGAAAGTTATGCTTAAGTATTTAGCTAATTAAAC 2920
Qy 294 GACCGAAGCATTTGATCTTGA---TCATCATTTAATCTAGACCTCGATATCA 344
Db 2921 TTAATGAGAGACTTACGCAATAGAAAGCTGTTGATTAAGATTTTCTGTAGATGC 2980
Qy 345 TTCAACAAGATATCTGAACCAAGACATCTCGAATGAGAGAAAAATTTCTTTTACA 404
Db 2981 CTACCAATTTCTAGAGATGACGTTACCAACATCTGCGCAACAGAAATTTAGATTTC 3040
Qy 405 TTCTCGGCAATTAACCAAGTGAAGACCGGATTAAGTTTATCGGCAAGATTAGTGC 464
Db 3041 TTGCGACAGAATTGAAGAAAGTCCCAAAACAGGGCTGGCTCCCTCGGCAAGTTTACAC 3100
Qy 465 AGTTGTCGCAAGTTTATATCCATTTTATCC-----CAATGTATCAGTAC 515
Db 3101 AGTTTAACTACGACTTTGCGCTCTCTTTTGTATCGAACCCTGAAAAATATAGACAA 3160
Qy 516 GAAATAAAGATATTTTGACAAAGTTGACAGATTTGACATTTTATGCCCCAAAAAGAT 575
Db 3161 ATATAGAGAGATTAATCATTAATTTAGACAAAGTTGCTCATGTGCAAGCTCAGGGTAAAT 3220
Qy 576 AGAATCGGGTTTGAATGTTGCACTGCAATTTATGCTGATTTATATGAAGATTTC 635
Db 3221 TGAAGCGGGTTGATGTAACGGCGGACGATATGATCTATCAGATATAGAAATTTCC 3280
Qy 636 GCCAGCTTTGATTAATGACGTTTCAAGTTCTTGAAGATGATCCTGAGAAAGTTCCAC 695
Db 3281 ACCCGATTAACTCTTAATTTGCGCAATTT-----GGAAGTCTTAATCGGCAATBA 3334
Qy 696 AGAATGAAAAAATTTGATTAAGATTAATCTGGAATTTCAACATGAAAGATGATCAATAC 755
Db 3335 ACTGGCGCATTTGGTTGATGAAGAAAGCTGGAATTTACGATTAAGTAACATTTACC 3394
Qy 756 ATACGGAATCAAGTTATTAATGAGTGAACGTCAGAGGGTCTCAGAAACCCCAATTTGT 815
Db 3395 TTGGGGATTAACTTTATGATGGCGATTTAAGAAATGTTTCAAGAAACAGTAAACTGGT 3454

```

QY 816 ATCAGAGTACTCCATGAGAAAAGGAAAGCCAGAAAGACTCTGTTGTATGACCA 875
 DB 3455 CCAGAAAGTAAAAATTTGGTATGATTCGCATATGCCAGAACTTGAAAATATATACAGA 3514
 QY 876 GCTTAATAGTCCCAATTTACAGTTTATGAAAGAAAT 911
 DB 3515 ACTCGATCATGCAAAATTTAGATTTATGATGAGACT 3550

RESULT 35

US-10-369-493-25849
 ; Sequence 25849, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 25849
 ; LENGTH: 1269
 ; TYPE: DNA
 ; ORGANISM: Schizosaccharomyces pombe
 US-10-369-493-25849

Query Match 6.7%; Score 86.6; DB 15; Length 1269;
 Best Local Similarity 56.5%; Pred. No. 1,88-11;
 Matches 205; Conservative 0; Mismatches 149; Indels 9; Gaps 2;

QY 499 CCCAATGTTATCAGTACGATTAAGATATTTTGCACACGTTGCACAGATTGACATTTGT 558
 DB 547 CTCAAAATGATGATTAACCAAGTCAATGTTGACATCTGGCACAATTTGCCACTGT 606
 QY 559 TATGCCAAAAAAGATAGATCTGGGTTGATGTTGCACTGCAATTTATGCTTGATT 618
 DB 607 TCTCACAAGAAAGTAAAGTGGTGTGATGTTGGCGCTGCCACTGGGGAAGTTGT 666
 QY 619 GTATATGAAGATTTCAGCCAGCTTTGATTAATGACGTGT-----TCAGTTTAGAA 672
 DB 667 ATTATAGAAAGTTGATCTTAATTAATTGAACAAGTGTGCTGCCATACGACGAACA 726
 QY 673 AGTATCTGAGAGTTCCCAACAGAGTTGAAAAATTGATGAAGTAACCTGGAAATTC 732
 DB 727 ATTAAGAACATTAATTTTTCACCGAATGAAAAAATGTTCTAAAAAATG---TCA 783
 QY 733 AAACATGAAGATGATACATACAGGAATCAAGTTATTAATGAGTACGTCAGAGGT 792
 DB 784 GAGTTGTTCATTTCAATTTGCTGCTACTTATTTGTTGTTAATGGGAGACGTAAGTGT 843
 QY 793 GGCTCAGAAACCCCAATTTGATCAAGTACTCCAAATGAAAAAGAAAAAGCCGAA 852
 DB 844 GGCAGTAGACACCGGGTATGTAATAAAGTACACCAATGGCAAAAGGAAATCCAGAG 903
 QY 853 GAA 855
 DB 904 GAA 906

RESULT 36

US-10-320-797-2218
 ; Sequence 2218, Application US/10320797
 ; Publication No. US20040014955A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Erosbkin, Alexey M.

APPLICANT: Zamudio, Carlos
 ; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
 ; FILE REFERENCE: 10182-021-999
 ; CURRENT APPLICATION NUMBER: US/10/320,797
 ; PRIOR FILING DATE: 2002-12-16
 ; PRIOR APPLICATION NUMBER: 60/341,261
 ; NUMBER OF SEQ ID NOS: 3361
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2218
 ; LENGTH: 1506
 ; TYPE: DNA
 ; ORGANISM: Cryptococcus neoformans
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1285)..(1285)
 ; OTHER INFORMATION: n=g, a, t or c
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1305)..(1305)
 ; OTHER INFORMATION: n=g, a, t or c
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1311)..(1311)
 ; OTHER INFORMATION: n=g, a, t or c
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1313)..(1313)
 ; OTHER INFORMATION: n=g, a, t or c
 US-10-320-797-2218

Query Match 5.2%; Score 67.6; DB 16; Length 1506;
 Best Local Similarity 48.8%; Pred. No. 1,5e-06;
 Matches 209; Conservative 0; Mismatches 213; Indels 6; Gaps 1;

QY 718 AGTAACGGAATTAACATGAAGAAGTACATTACATACGGAATCAAGTTATTAATG 777
 DB 1036 AGCAATAGGATTAACAAATCTATCCCTTCCGACTCCCAAACTTTGCGTCTTACTC 1095
 QY 778 GGTGACGTCAAGGTGCTCAGAAACCCAAATTTGATCAAGTACTCAATGAAA 837
 DB 1096 GCCGACGTTTGTGGGTACTGACACGCACTTTGTCTCTCGTCTCAAGTGAAGA 1155
 QY 838 AAGGAAGCCAGAAAGAGCTGTTGTGTATACCAAGCTTAATAGTCCAAATTTACAG 897
 DB 1156 AATTAAGTAAAGAAAGCCGATGAAGTGGGCGAGCTGATTAATGGAACAGGGCC 1215
 QY 898 TTATGAAGAAATGAGGGAATGCTGAAAAAATACGACTCAGACCCAGAGACTTATATT 957
 DB 1216 TTGGGTAGGCTGGAAGATATGCTTTGAACCTCCCTACACCCGAGCCCTTACGC 1275
 QY 958 AAAGATTAGATCTTCTGTTGAGCCCTTTGACTGTTGCGATTAAGAACATCAGAAAAGG 1017
 DB 1276 ACACCTC-----CATTTCTTCACCGCGCTGCTTNTGCCCTCATCCGCGCTCG 1329
 QY 1018 TTACAAGCATTAACAAAAATCAGAGTTCAATTGAACTGATGTCGAAACCCAGTTG 1077
 DB 1330 TTGGGTAAATGCTCCGACTCTGGGGCTGACCAATGCAACAAAGAAAGAAATTTG 1389
 QY 1078 TTGACCGTTGTCAAGAGATTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGA 1137
 DB 1390 CTGATGATCGGACGAAAGTCAAGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1449
 QY 1138 TACGATGC 1145
 DB 1450 TACGATGC 1457

RESULT 37

US-10-128-714-2406
 ; Sequence 2406, Application US/10128714
 ; Publication No. US20030119013A1

```

; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; PRIOR FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2406
; LENGTH: 1356
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-2406

```

```

Query Match 5.1%; Score 66; DB 15; Length 1356;
Best Local Similarity 45.7%; Pred. No. 3.7e-06;
Matches 547; Conservative 0; Mismatches 595; Indels 54; Gaps 7;

```

```

QY 6 AAAAGCATTAGTGCACCTGGAAAAAGCATTCTTCTGCTGCTGATATTTGCTTTGAGCC 65
DB 24 AACGGCTTGTCCCGCGCGGCAAGTCTCTCTACCTGAGGGGTATCTGGTCTCTGACCG 83
QY 66 AATTATGATGCTATGTAGACAGCATTTGTCTACGATGATGATGATTAACCAAA 125
DB 84 CAGCTTACACTGGGAGCTGTGTTCGCTTACGCGCAAGATCCATCGAAGGCAATGTTC 143
QY 126 AGGAACCACTTTGAAAGATCTAGAAATCAAAATTTCTTCAACCCCAATTTGCAAGAGA 185
DB 144 TGGAGACAAAGAAAGCAAGGACGATCTGTACACTCCCAAGTTGCTGATGGGT 203
QY 186 ATGGGAATATCATATCATCAATAATACAGAAAGCCAGAGAGTTCACTGACGATPAA 245
DB 204 ATGGAGTATAGCATCAAAAGATCGAGACGAGAGTGAACCCATCCAGGCCAGCC---AA 260
QY 246 TCCATTTTGAAGGCACTATATTCATCGTTTGTAGTTATAT-----TCAACGAGC 297
DB 261 CCGTTTGTGAGACCTCTCTGAACCTTCTGACTTACATCACTATGTGCGGACTC 320
QY 298 GAAGCATTTGATCTTGAATATCATATTAATCAAGACCTGATATCAATTCAGAAAGAT 357
DB 321 GAAGGATTTGGGGTATATGATTAATCTATCTGCGGCAACAATATCACTCCGAGAC 380
QY 358 ACTGAAACCAAGACATCTCGAATGAG-----AAAAAATTTCTTTTCAATTCCTGTC 413
DB 381 GGGCTTCTCCAGGCTTCGGGACTCCGGTCTCAAGAGATTCGTGACCTTTGTGTTGG 440
QY 414 CATTACGAGTGGAAAGACCGGATAGTTCAATCGGAGAGATTAGTGCAGTTGTTGC 473
DB 441 CTTTCAGAGGACACAGAGAGGCTTGGGCTTTCAAGCCCTTTGCTCACTGCTGTGT 500
QY 474 CACAAGTTTATATCCATTTTA-----TCCCAATGTTTATCACTAGATTAAGATAT 527
DB 501 GTGTCTCTGTATCAACGCTATATGACAAACCGAGAGATCTGGTCCAGGCGGACAA 560
QY 528 TTGCAACAAGTTGCAAGATTTGATTTATGCCCCAAAAAAGATAGAGATCTGGGTT 587
DB 561 GCTTCAAAATCTGGCCAGGCGGCCCACTGCGCTGCCAGGATTAAGTCTGGGCTT 620

```

```

QY 588 TGATGTGCAACTGCAATTTATGTCTGATTTATATAGAGATT-----TCAGCAGCTT 643
DB 621 GATGTGCAAGCTGCAATTTATGCGCTCTGTCTTAAGAGCTTCTCCCTGATTTCT 680
QY 644 TGATTAATGACGTGTTTACGTTTATAGAAAGTATCTTGAGAGTTCCCAAGATTTGA 703
DB 681 CGATTCAGTGGGTGACCTGTGTCTCCAGCTTCAAGAGCGGTGTGTTCCGATCTGAGA 740
QY 704 AAAAATTGATTGAAGTA-----ACTGGAATTCAAACATGAAGATGTATCAATACC 755
DB 741 GAGCGCCGACCCCGAGATCCGTGGGACCCGAATGTCTAGACTTCGCGATGAAGCTCC 800
QY 756 ATACGAATCAAGTTATTAATGAGTGAAGTCAAGGCTGCTCAGAAAACACCAATTTGT 815
DB 801 CCGGGAATGAATGATCTCTGTGTGAGCTGATGATGTGTTCCGAGATCTCATCATGT 860
QY 816 ATCAAGATCTCAATGAAAAAGAAAAGCCAGAAAGAAAGCTTGTGTATGACCA 875
DB 861 GAGAAAGTTTGTGAGTGGCGGAAACAGAAATCAGAAAGAACCGATATGCTCTGGGCGC 920
QY 876 GCTTAATAGTCCCAATTTACGTTTATGAGGAATGAGGAAATGGGTGAATAATGCA 935
DB 921 TGTCAATCGAACAAGAGACTTCCCTGGAATCTGACGCTTGACAGAGCCCGGA 980
QY 936 CTCAGACCCAGACTTATATTAAGAGTTAGATCATTTCTGTTGAGCCTTTGACTGTGC 995
DB 981 CGAATATCTCTAGTATCTTGAATAATGTCCGACTATATTCAGGCTCTG----- 1032
QY 996 GATTAGACATCAAGAAAGGTTTACAGAGATTAAACAAATAACAGAGTTTCAATTGA 1055
DB 1033 -----CGTAACCAATCCGTTCCATGACTCAAAAGTCGATGCCAATCGA 1079
QY 1056 ACCTGATGTCAAACCCAGTTGTTGACCCGTTCAGAGAAATTCCTGCTGTGTG 1115
DB 1080 GCGCGGCTCAACCAAGCTTACTGACGCTCTGTCCGAGCTGAGGAGTGTATCTGG 1139
QY 1116 TGTGTTCCAGAGTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1171
DB 1140 TGTGTTCCAGAGGAGGAGGAGGCTTACGACGCTATGTCTCTCATCAAGCAATC 1195

```

```

RESULT 38
US-10-032-585-3023/c
; Sequence 3023, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3023
; LENGTH: 65
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-3023

```

```

Query Match 5.0%; Score 65; DB 15; Length 65;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ATGCAAAAGCATTAGTGCACCTGGAAGAAAGCATTCTGCTGCTGATATTTGCTTCT 60
DB 65 ATGTCAAAAGCATTAGTGCACCTGGAAGAAAGCATTCTGCTGCTGATATTTGCTTCT 6
QY 61 GAGCC 65
|||||

```


Db 5 GAGCC 1

RESULT 39
US-10-128-714-7406
; Sequence 7406, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastian M
; TITLE OF INVENTION: Identification of Essential Genes in *Aspergillus fumigatus* and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT FILING DATE: US/10/128,714
; PRIOR FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7406
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: *Aspergillus fumigatus*
US-10-128-714-7406

Query Match 4.8%; Score 62.6; DB 15; Length 1455;
Best Local Similarity 46.7%; Pred. No. 2.8e-05;
Matches 375; Conservative 0; Mismatches 389; Indels 39; Gaps 4;

Db 387 AAAAAATTCTTTTACCATTCGTCGCCATTAACGAAGTGAAAAGACCGATTAGTTC 446
513 AAGCAGATTGCGAGACTTTGGTGTTCGCTTAGAGGCACACAAGAAGCGCTGGGCTC 572
QY ATGCGCAGAGATTAGTGTAGTGTGTCGCAAGTTTATTATCCCA-----TTTATCCC 500
Db 573 TTCAACCGCCCTTGTCACCTGCCCCGTGTCCTCTCGTCATCCAACCTACTATGCAACC 632
QY CAATGTTATTCAGTAGAAATTAAGATATTTTGACACAACGTTGCACAGATTGACATTGTTA 560
Db 633 CGACCATCTCGGTCCAGGCGCGGACCAAGCTTCAATCTGGCCCAAGGCGGCCACTGGC 692
QY TGCCCAAAAAAAAAAGATAGGATCTGGGTTGATGTTGCAACTGCAATTATATGCTGATTTG 620
Db 693 TGCGCAGGGTAAAGTCGGGTCGGGCTTCGATGTTGCACTGCACATTATCGGCTCCTGCT 752
QY ATATGAAGATTTCAGCCACGCTTTATATAATGACGTGTTCA----GGTTCTAGAAAGTG 676
Db 753 CTACGACAGCTTCTCCCCCTCGATTCTGGAATCAAGTGGAGTACGCTGGTCTCCAGGCTT 812
QY ATCTGAGAAAGTTTCCCAACAGAGTTGAAAATAATGATGAAAGTAAGTGG-----A 728
Db 813 CGAAGAGCGGTTGTTCCGATGTGTAAGAGACCCGACCAAGCATCCGTGGACACCGA 872
QY ATTCAAACATGAAGATGTACATTACATACGGAATCAAGTATTATTAATGAGTGAAGTCAA 788
Db 873 ATGTCTAAGCTTCGGCATGAGAGCTCCCCGCGGAATGCAAAATGCTCTGTGACGTGA 932
QY GGCTGGCTCAGAAACACCAATTGGTATCACAGACTCTCCAAATGAAAAAGAAAAAGCC 848
933 ATGTGTTGCGCAGACTCCATCCATGCTGTAAGAAAGGTTTTTGAGTGGCGAAACAGAAATCA 992

[illegible]

```

RESULT 40
US-10-128-714-1406
; Sequence 1406, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eiroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-9939
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1406
; LENGTH: 1509
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-1406

Query Match      4.8%; Score 62.6; DB 15; Length 1509;
Best Local Similarity 46.7%; Pred.No. 2.9e-05;
Matches 375; Conservative 0; Mismatches 389; Indels 39; Gaps 4;

QY      AAAAACATTTCTTACCATCTCTCGTCGCATACCGAAGTGGAAAGAACCGGATTAGGTTC 446
DB      AAGCAGATCTCGAGACTTTGGTGTTCGCTTCAGAGAGCACACAAGAGCGCTGGGGTTC 626
QY      ATCGGACAGGATTAGTGTCACTGTGTGCACAAGTTTATATCCCA-----TTTATATCCC 500
DB      TTCAGCGGCTTGTGTCACTGCGCTGTGTGTCTCTCTGTATATCAACGTACTATAGCAACC 686
QY      CAATGTATATCAATACGAATATTAATTTTGGACACAGCTTGCACAGATTTGACACTTGTTA 560

```


Db 687 CGACGATCTGGTCCAGGCCGCGAACAAGCTTCACAACTGTGCCCCAGCGGCCCACTGCGC 746
QY 561 TGCCCAAAAAAGATAGAGATCTGGGTTTGATGTTGCAACTGCAATTATATGCTGATGT 620
Db 747 TGCGCAGGGTAAGTCGGGTCCGGCTTCGATGTTGCACTGCACTTACGGCTCCTGCT 806
QY 621 ATATAGAAAGATTTCAGCCAGCTTTGATTAATGACGTGTTTCA---GGTTCAGAAAGTG 676
Db 807 CTACAGACGCTTCTCCCCCTGATTTCTGAATCAAGTGGGTGACCGTGGTCTCCAGGCTT 866
QY 677 ATCCTGAGAGTTCCTCCACAGAGTTGAAAAATTGATTGAAAGTAACTGGG-----A 728
Db 867 CGAAGACGGGTGTTCCGAGATGTAGAGAGCCGACCCCGACATCCGTGGACACCGA 926
QY 729 ATTCAACATGAAGATGATCATTAACCAATACGAATCAAGTATTAATATGAGTGAAGTCA 788
Db 927 ATGCTTAGACTTGGCATGAAAGTCTCCCGCGAATGCAATGGTCTCTGTGACGTGCA 986
QY 789 GGGTGGCTCAGAAAACCCCAAATTGATATCAAGATCTCCATGATAAAAAGAAAAGCC 848
Db 987 ATGTGTTCCGACATCTCATCATGTGAGAAAGSTTTTGAGTGGCGGAAACAGAAATCA 1046
QY 849 AGAAGAAAGCTCTGTGTGATGACAGCTTAATATGTCCTCAATTACAGTTTATGAAGA 908
Db 1047 GAAGGAAGCCGATATGCTCTGGGGCGCTGTGCAATCGAACACGAGAGACTTCGCTGGA 1106
QY 909 ATTGAGGGAATGCGTGAAAAATAGACTCAGACCCAGAGACTTATATTAAGAATTAGA 968
Db 1107 ACTGAGACGCTTGGCACHAGGCCGAGACGAACATCTCTCAGTGACTTTGAAAAATGTCG 1166
QY 969 TCATTTCTGTGAGCCTTGACTGTGCAATTAGAACATCAGAAAAGGGTTACAGCAAT 1028
Db 1167 CACCTATATTCAGCGCTCG-----CGTAACCAATCCGTTCCAT 1205
QY 1029 AACACAAAAATCAGAGGTTCCAAATTGAACCTGATGTCCAAACCAAGTTGTGAACGTTG 1088
Db 1206 GACTCAAAAGTCGATGTCCCAATCGAGCCGCGCTCCAAACCGAGCTACTTGAAGCTCT 1265
QY 1089 TCAAGAGATTCCTGTTGTGTGTTGTTGTTGTTCCAGTGTGCTGTGGAATACGATGCAAT 1148
Db 1266 GTCCGAGCTGAGGGGTGTCACTGGGTGTGTGTTCCAGAGCAGGGGCTACGAGCCAT 1325
QY 1149 AGCTGATTAATGTTGAAAAATC 1171
Db 1326 TGTGCTCCTCATCCAAGACAATC 1348

Search completed: January 26, 2005, 07:56:03
Job time : 566 secs

This Page Blank (uspto)